VOL. 27, No. 6 November, 1920

THE PSYCHOLOGICAL REVIEW

DO WE THINK IN WORDS?

BEHAVIORIST VS. INTROSPECTIVE CONCEPTIONS

BY ARTHUR S. OTIS

Stanford University

1. *Purpose of the Discussion.*—The following discussion takes its departure from the reading of Dr. John B. Watson's 'Psychology from the Standpoint of a Behaviorist.'

It is the purpose of the writer to discuss certain hypotheses which are put forth with seeming conviction in the text but which are believed by the writer to be false.

To discuss the whole subject of the Behaviorist point of view, in relation to the more generally accepted points of view in psychology would be quite impossible in the scope of this article. That a text in psychology should be written in which the author not only purposefully avoids the mention of such concepts as perception, ideation, association of ideas, consciousness, attention, will, etc., but even goes so far as to claim that these concepts are useless for purposes of psychology, is of course quite a source of wonder. The indispensability of the concepts avoided by Behaviorist psychology and of the use of introspection will be apparent, we believe, from the discussion of but one 'assumption' which it makes. We shall confine this article to the discussion of this assumption.

The hypothesis referred to is 'the point of view that has been advocated throughout the text, namely, that thought is the action of language mechanisms' (p. 316). The meaning of the expression, language mechanisms, is carefully defined by the author as referring to any of those muscles

of the body which actuate to produce words whether spoken, written, or gesticulated (as by deaf mutes). The meaning of the word, thought, as used in this hypothesis is not explicitly stated, but may be inferred with confidence from various passages which we shall quote[1] and is here taken to be the same as the meaning of thought when used by those who are not Behaviorists.

To be sure, the author states in the preface that "the terms thinking and memory have been carefully redefined in conformity with Behaviorist psychology." On page 14 we find in italics the expression: '"thinking," by which we mean subvocal talking.' This may constitute the re-definition, but if so it obviously begs the question which we are discussing; namely whether 'thinking' as ordinarily understood does consist of subvocal talking. We shall therefore leave this re-definition out of account.

It will be realized that the adjustment of an individual to his environment may involve acts requiring mental activity of all degrees of consciousness, from the most automatic habitual or instinctive acts requiring little or no consciousness, such as moving the eyes toward an object it is desired to see, to the solving of problems requiring the

[1] "A man may sit motionless at his desk with pen in hand and paper before him. In popular parlance we may say he is idle or 'thinking,' but our *assumption* is that his muscles are really as active and possibly more active than if he were playing tennis. But what muscles? Those muscles which have been trained to act when he is in such a situation, his laryngeal, tongue, and speech muscles generally" (p. 15).

"We manipulate vocally" (when trying to think of the name of a familiar person) "by running over the names beginning with each succeeding letter of the alphabet, or by saying 'black hair,' 'blue eyes,' 'six feet tall,' and the like" (p. 305).

"The explicit and implicit language habits are formed along with the explicit bodily habits and are bound up with them and become a part of every total unitary action system that the human organism forms. . . . They are present in the simplest types of adjustment that he makes. We can see the functioning of language habits only slightly in certain activities, as, for example, in swimming, tapping on the table with a pencil, while in certain other types they form an integral part . . ." (p. 309).

"Our view is that overt language develops under social training. It is thus absorbed into and becomes a part of every total integration of the individual. Hence when he is making adjustments in the absence of other like beings language remains as part of the process" (p. 323).

". . . the maiden thinks of her lover in words . . . . . the beautiful thoughts of the idealist for mankind as a whole or of the mother for her child . . . are couched in words or their equivalent" (p. 325).

most concentrated mental effort. It is hardly conceivable that the Behaviorist would claim that *all* such adjustment involves language mechanisms, as the passages quoted would imply if taken literally. To simplify matters, however, we will limit our discussion to that type of adjustment ordinarily referred to as 'thinking,' namely, those mental processes of the problem-solving sort which require some degree of conscious mental effort, since these are open to introspective investigation. That even these processes of adjustment do not necessarily require language we shall attempt to show by appeal to logic and common experience, omitting arguments *ad hominem*.

2. *Examples of Thinking.*—Let us consider one or two simple cases of problem solving and subject them to critical psychological analysis in order to determine whether they involve language.

Suppose I have unfolded a new map and am attempting to fold it again as it was. I have no complete habit, not having folded a map exactly like this before. Let us see what happens. Surely there is a better way than to let someone watch me and report his inferences. He would merely see me look at the map and, let us say, try to fold it in one way but fail and then try another way and succeed. He might infer that my method was the so-called 'trial and error' or 'perseverance' method. Or if my lips have moved he might infer that I arrived at the solution of the problem by means of the action of the muscles of my lips and other speech organs. This appears to be the method of the Behaviorist.

Let me introspect and report from direct observation what happened from the point of view of one looking on from the inside. The writer does not wish to be misunderstood as assuming that introspection is infallible. One's testimony is not infallible even when he observes with his own eyes an incident which transpires directly before him. Relatively speaking however, introspection is far more direct and reliable than inference based upon observation from the outside.

On introspection I report as follows regarding my action with the map. More or less mechanically, as we say, that is, while thinking about what I had seen on the map, I began to fold the map along one of the creases. After a moment I became aware that the map was not falling into its accustomed folds. I then became aware of the need of finding the correct way to fold the map and I unfolded it in order to begin again. I recalled from previous experience that the crease on which the first fold must be made is one which runs entirely across the paper. I therefore looked for such a fold and on finding it folded the paper on it and repeated the process until the map was entirely folded up, making no further error.

Now this adjustment which I have made to the problem of folding the map was 'thinking,' alike in the popular usage and that of the psychologist. The Behaviorist claims that thinking is the action of language mechanisms. Let us go over this example of thinking again and examine it very minutely to see whether there is any necessary connection between language and the solving of the problem.

First of all, how do I become aware that the map is not falling into its accustomed folds? If I go slightly back of this awareness I note a feeling of contradiction between a subconsciously expected feeling of flatness and the experienced feeling of bulginess. This contradiction, we may say, caused me to become aware of the improper folding of the map—caused the shift of my attention from the thoughts of what I had seen on the map to the matter of folding the map.[1] How did I then become aware of the need of finding the correct way of folding the map? The experience suggests no other explanation than merely to say that the idea of contradiction 'called up' or 'suggested' the idea of need. This idea in turn called up the idea of beginning again. We may explain this process by saying that it was probably

[1] That a subconscious awareness of contradiction may give rise to an idea of need, together with an affective state which effects a shift of attention (clear awareness) to the need, is a matter of so frequent observation in structural psychology as to be considered a scientific fact. Such a fact, however, is of course quite unthought of in Behaviorist psychology, being wholly outside of its scope.

the result of a previously formed habit. One has learned in such cases that it is best to begin again. When the idea of unfolding the map again has come to occupy more or less of the whole of consciousness, 'the thought takes form in action.' Behaviorist psychology concedes such a phenomenon, so we need not attempt to explain it. Having unfolded the map I recalled previous experiences regarding the folding of large sheets of paper. We will say that the perception of the paper before me and the idea of need of folding, together served to bring forth from my memory store those ideas which came to my mind. These together with the perceptions of the map during the process of folding served to educe that train of ideas which guides the folding to a successful termination.

Now what is the material of all this mental activity? What do these ideas consist of? They consist of images, visual, tactual, kinæsthetic, etc., of maps, and of certain aspects of these images such as creases, folding movements, flatness, bulginess, etc. They have nothing to do with language, necessarily. The idea of flatness is tactual or visual or both, the idea of a folding movement may be visual or kinæsthetic or both. The idea of the length or direction of a crease is visual or kinæsthetic or both. Possibly other types of imagery enter to a slight extent. But no language need be involved.

Let us now consider another type of thinking. Let the reader ask himself why it is more difficult to play a game of chess blindfolded than with the chess board visible before him. Obviously the answer is that the perception of the relative positions of the chess men is a great aid to the mental manipulation which constitutes the basis of the study of moves. Moreover, anyone who has played chess or checkers will immediately appreciate the aid that would be derived from actually making the trial moves that are contemplated, in more clearly appreciating the relations that such moves would introduce. If the thinking were done by means of subvocal language it would seem that seeing or not seeing the chess board would make no difference. The obvious

answer is that the thinking is done by means of the perceptions of the board and men as they are, the mental imaging of the movement of the men into new positions and the appreciation of the spacial and temporal relations between the pieces and their possible moves as introduced after the mental manipulation. No language whatever is required. As we shall show, a person may indeed talk to himself while contemplating moves, but this activity is entirely secondary and supplementary.

3. *The Material of Thought.*—Thinking, as an adjustment of the individual to his environment, as the solving of problems, consists of the evolving of new ideas, concepts, or meanings, from old. This is accomplished by recombination of the elements of the old into new patterns. By ideas, concepts, and meanings are meant image patterns, whether they be of words, objects seen, sounds heard, things felt, tasted, sensed in any manner whatsoever, or any quality, attribute, or aspect of such image patterns as may be conceived separately by abstraction, such as shape, color, surface, volume, extent, duration, intensity, symmetry, movement, similarity, difference, causality, symbolism, abstractness or affective quality; or of whatever degree of clearness or attenuation or incipiency the images or image aspects may be. We may think, therefore; that is, we may evolve new ideas, concepts, meanings, in terms of image patterns of any kind whatsoever, or of the consciousness (idea) of any relationship whatsoever between these image patterns.

For example, I am thinking when I am effortfully engaged in composing a piece of music. I sit at the piano with music paper at my side. My mind is occupied with perceptions and images of tones, tone combinations, tone sequences, tone relationships, tone emotional effects, tone symbols (dots on paper) the making of these symbols, etc.

My effort consists in the maintenance of my attention to the work, the calling up of sequences of tone images,[1] the

[1] Strictly speaking I adopt the mind set that will result in the calling up of tone images, or that is calculated to do so. (Sometimes I may succeed better than at other times.) We cannot call up an image necessarily at will. Generally it is a case of taking a certain mental attitude ordinarily called 'trying to think' which usually results in the recall of the idea desired.

comparison of these, the appraisal of their respective æsthetic values, the choice of one or another, the calling up of the proper symbols of notation in which to write down the musical ideas, and the writing of these. No language is involved in any of this thinking (except perhaps a final translation of the results of thought into symbols). In this case also I may compose without the piano; but this is more difficult, since I am compelled to make my judgments upon images only, whereas with the piano I may employ the perceptions of the tones themselves in my judgments. If my musical thinking were all done by means of the action of speech muscles we do not see that it would make any difference whether the piano were struck or not.

Similarly, one is thinking when he is creating a new architectural design, or a drawing or painting or statue or stage setting, or conceiving of a new dance movement or inventing a new mechanical contrivance or playing tennis or searching for the cause of engine trouble. The material of one's thoughts in all these cases is in the form of images, which need be only visual, auditory, tactual, kinæsthetic, may be, in fact, of any kind whatever according to the requirements or to one's ability to call forth such images. As we have said, one may do any amount of talking to oneself while thinking—which is merely putting one's thoughts in words after they are thought—but the talking is not the thinking. It is supplementary to it in exactly the same way that describing a landscape is supplementary to seeing it.

Thinking may be called the controlled association of ideas, in contradistinction to the free association of ideas. In the free association of ideas, by which we refer to what is ordinarily called day dreaming or revery, ideas follow one another in a more or less unguided manner, yet in a fairly rational way as compared with the incongruous manner of idea sequence sometimes experienced in dreams. Doubtless there is some sort of control even in 'free association' though it may be the general interest in the subject of thought or the control occasioned by thought habits. However, in what we have called controlled association of ideas characterizing

thinking, the ideas are guided in their sequence by some conscious aim, *e.g.*, a problem to be solved. Irrelevant ideas are discarded (attended from), relevant ideas are attended to. That which does the controlling is often also in the form of a definite idea. This is best illustrated when one is given two digits written one above the other: if told to add them, under the influence of this guiding idea they call forth their sum; if told to subtract one from the other, under the control of this guiding idea they suggest their difference. The same stimuli give rise to either one or another idea according to the nature of an additional and controlling idea.

Similarly, we may have occasion to think of the *opposite* of a given concept, or of a subordinate, or the superordinate or the symbol of a given concept. In any case one idea calls up a second under the guidance of a third.

We may not only occupy our minds with ideas of the color, size, shape, etc., of objects, as referred to above, but we may *compare* two objects as to color, size, shape, weight, motion, acceleration, symmetry, etc., and *judge* which is the best suited to our needs. Of two individuals we may compare the good looks, cordiality, sincerity, hospitality, integrity, adaptability, intelligence, etc., as conceived in ideas of conduct, feelings, appearance, facial expression, and of the many circumstances under which the impressions were gained. All these mental activities dealing with ideas as material—their association, recall, generalization, abstraction, comparison, judgment, etc., are elements in adaptation, yet they may be experienced or accomplished quite independently of words. The idea of a color is not a word. The idea of one color being more intense than another need not have anything to do with language. The choice of this or that color for an æsthetic purpose does not require language, nor does the act which the choice calls forth. Yet all this is adaptation.

4. *Words may be the Material of Thought.*—As has been suggested throughout the discussion, words may be the material of thought. The place of words in the range of material of thought may be stated as follows. The material

of thought, as explained below, begins with perceptions; then come images resembling perceptions, then more and more attenuated images or aspects of images singled out by abstraction, and finally symbols. By symbol is meant any concept which is used in place of another. The best illustration of thinking in symbols is in the case of the number symbol system used in arithmetic and algebra. The idea of eight (not in the word but the number: ********) is represented by the symbol: 8. The idea of seven (seven things: *******) is symbolized by the figure: 7. Now if we have the problem of finding the sum of these numbers (******** and *******) we may do so by translating them into their respective symbols and give our attention to the symbols only. Having previously formed an association between the symbols, 7 and 8, and the symbol of their sum: 15, the symbol 15 is called up when the symbols, 7 and 8, and the guiding idea of summation are in mind. We may then proceed to make other arithmetical computations in terms of number symbols only, letting these call up the number idea (****—) when needed. Similarly in algebra we may let x represent one number with which the problem deals and let y represent another number, etc., and then by means of habits established in connection with these symbols we may do thinking of a simple type in lieu of what would be far more difficult if done with the original concepts of number. This type of thinking is exemplified in the following algebraic reasoning:

If $x^2 - y^2 = z$ then $(x - y)(x + y) = z$.

There are of course many kinds of symbols. In addition to the number symbols just mentioned there are the symbols of operation upon numbers such as those of addition, multiplication, integration, involution, etc., there are the symbols of musical notation, symbols of punctuation (?, !, ", *, -), symbols on maps representing roads, trees, buildings, bridges, tunnels, etc. (an engineer can think very effectively in these symbols). There are even symbolic facial expressions used by actors to portray emotions which off the stage would not be expected to produce those expressions. A skull and cross bones symbolizes danger. The flag symbolizes country,

etc. Last, and most important, of course, words and sentences symbolize thought of every description. Occasionally we feel that we have experienced some thought or sensation or feeling which cannot be expressed in words. But in general all ordinary thoughts and feelings can be represented by some word or sentence.

We see therefore that language constitutes only one of the various kinds of symbolization, and symbols constitute only one type of material of thought.

Words are themselves the material of thought under many circumstances. Whenever we have to communicate thoughts to another or learn the thoughts of another through language we have to deal with words. By far the greatest use of language of course is in the calling up of language symbols to represent meanings or the calling up of meanings represented by language. Occasionally however we may think in terms of language almost exclusively, as when dealing with the rhyme and rhythm of poetry. In the case of syllogistic reasoning we may be truly said to be thinking in words, when the expression "All *A* is *B* and all *B* is *C*" calls up the language idea: "All *A* is *C*," or when part of a sentence suggests the rest as "All is not gold that ——."

5. *Language the Symbolization of Meaning.*—We have attempted throughout this discussion to distinguish clearly between a *meaning* and the language by which it is symbolized. We cite the following illustrations to bring out this distinction still more clearly.

If one says: "I saw John Jones on the street this morning" the hearer will get the meaning of the sentence at once. "Getting the meaning" means to the ordinary person getting an image, more or less faint perhaps, of the speaker looking at Jones on the street. But let us take another sentence. Here is one in which the meaning of a new (coined) word is stated. Every word in the statement of the definition except the new one is perfectly intelligible and familiar and the statement is a perfectly logical and meaningful one, yet we are confident that the reader will not get the meaning from the language on first reading. This is the sentence: "Let

us define the word, incration, as meaning the increase in the number of feet per second per second by which the motion of a body is accelerated." Anyone who has gotten the meaning of this sentence clearly should be able to point out immediately the error in the following statement, which if correct would follow as a corollary to the above definition: "The unit of incration is one foot per second per second." (The correction is indicated in a footnote.) If the reader is unable to point out the error it is merely because he has not gotten the meaning of the definition, which is something quite apart from the words by which it is symbolized and consists of images either of the motion of a body or of the path of its motion. Without such images we are confident the meaning can in no way be appreciated.[1]

As has been said, the utterance of sentences or of parts of sentences or of analogous statements often *helps to bring out* the meaning, that is, helps to call up the imagery necessary to build up the meaning, or helps to fix the meaning in mind by symbolizing it after it is appreciated. But the meaning may exist entirely independent and apart from any utterance, either overt or implicit.

One often hears the expression from pupils in school: "I know but I can't tell." This is generally a simple case of having a meaning or idea without the ability to symbolize it in language.

Moreover, as has been stated, even an adult may have experienced perceptions, ideas, or feelings which he will declare cannot be expressed in words. Even if they could,

[1] The correct statement is: The unit of incration is one foot per second per second per second.

A reasonable comprehension of the meanings of these statements may be built up with the help of the following leading statements. The *rate* of motion of a body is the number of feet per second which it moves. The unit of rate is one foot per second. The *acceleration* of a moving body is the increase in its rate, that is, the increase in the number of feet per second which it moves in succeeding seconds. The unit of acceleration is one foot per second per second, that is, one foot per second every second. And again, the *incration* of a moving body is the increase in its acceleration, that is, the increase (from second to second) in the number of feet per second by which its rate is increased, or in other words, it is the number of feet per second per second by which the motion of the body is accelerated. The unit of incration is the unit of acceleration every second, that is, it is one foot per second per second per second.

his inability to do so testifies to the independence of the thoughts from the language by which it would be symbolized. We say of our memory of a sunset that it was 'indescribable' and 'would have to be seen to be appreciated,' which is entirely true. We say, 'Words fail me' regarding the expression of our thoughts of past emotional experiences. The emotional experiences of love, hate, fear, anger, etc., may be the material of thought just as well as the experience of perceiving the color blue. Even our best attempts to express our thoughts of experiences in words often fail to carry full meaning to the hearer unless he has had a similar experience. Thus, one may say that an experience was like that of a sudden drop in an elevator or like flying in an aëroplane, but unless the hearer has had an analogous sensation or experience the expression is devoid of essential meaning to him. What can the expression, 'like being struck by lightning' or 'like finding oneself caught under the ice' mean to one who has not had the experience in comparison to what it means to one who has had the experience! Yet the language is identical in the two cases. One and the same expression from the lips of an oratorical person may convey a meaning to one hearer which will call forth tears, a meaning to another hearer which will call forth anger, and a meaning to still another hearer which will call forth laughter. Such a phenomenon would be of course entirely impossible if there were but one meaning to the expression, a meaning inherent in the language itself. It is a platitude that the meaning of language is something which is brought to it from the experience of the hearer, that it does not reside in the language.

6. *The Genesis of Language.*—Not only may we have thoughts for which we cannot think of the existing appropriate language, but we may often have an idea for which there is no corresponding word or phrase. Indeed language is built up by the coining of new words and phrases that are needed to symbolize new thoughts for which no corresponding language exists. The word, automobile, for example, did not come into existence until after there had been made,

or at least conceived, a machine which would move itself, and which needed a name. Similarly the expression, 'carry on,' came into use in response to the need of a name for an action which was well comprehended but for which no convenient language equivalent existed. Ideas originate first; afterward they are named—symbolized in language.

Perhaps the clearest example of the temporal relation between the genesis of ideas and their symbolization is the case of the naming of persons. First the child, which we perceive, is born. Afterward it is given a name. Why do we give a child a name? It is, of course, for the reason that to refer to it always by description would be cumbersome and inaccurate. When we think of a person deliberately we think of his form and features, his speech, manner, expression, etc. When we think of him more fleetingly, however, our imagery becomes attenuated, even perhaps to the extent exemplified by the representation of Roosevelt by merely a pair of glasses and a row of teeth. But for the purpose of one person conveying the thought of an individual to another, such attenuated imagery is inexpedient. We therefore symbolize the whole picture or idea of the individual by a single word (a name). The name then, in cases of rapid thinking, may nearly take the place of the concept of the individual as imaged. The name, however, still does carry with it something of the original imagery. The idea of 'James' to one person carries with it something which characterizes his brother, James. The idea, James, to another person, carries with it something which characterizes his uncle, James, a different individual. This additional something is needed to constitute the difference in meaning between 'James' for the one person and 'James' for the other. To a third person who knows no one by the name of James, the idea is merely a word, known to refer to some individual.

On the other hand the perception of a person whose name is not known does not carry with it any idea of a name, nor need any word come to the mind. When I see a man I do not think 'man.' When I think of a crowd of persons I do not think of a crowd of words! Language is as distinct from

the ideas it represents as the name of a person is distinct from the perception of the person himself.

An infant of course does not use language habits until a year or so after birth. Yet an infant can think—perceive, compare, judge, choose, decide, act upon decision, etc.—before language habits begin. The material of his thoughts is the sights, sounds, smell, tastes, feelings, etc., which he experiences throughout his waking state.

To illustrate the genesis of language habits we must go back to the early days of a child's life when it is just beginning those 'Abbreviated and short-circuited actions (which) become a necessity if it is to hold its own in that environment and make progress' (p. 319). The child's perception of its doll, its desire for the doll, its idea of searching for the doll when not in sight, its idea of creeping toward the doll when seen, its idea of reaching for the doll when within reach, its idea of grasping and the new ideas which arise from the manipulation of the doll—assuming these ideas sufficiently well fixed by habit that the actions have reached an 'abbreviated and short-circuited' stage and are purposeful and adaptive—these ideas constitute the beginnings of thought (conscious adjustment to the environment).

The stimulus, 'tata' (p. 320), cannot call up the concept, doll, before the concept, doll, is in existence. Nor can it create the idea. The child must have some idea of the doll, formed from perceptions of the doll itself, *before* the idea of 'tata' can be associated with it. Language in general bears just the same relation to thought in general that the idea, 'tata,' as a word, bears to the idea of the doll, as something seen, touched, etc.

Similarly the idea of number is formed before any symbol representing number can be associated with it. We may talk about number in the hearing of a child for months after it has begun the use of language, but until the child has by observation and comparison become conscious of the two-ness of its hands, of the two-ness of its feet, and of the two-ness of many other separate things, so that the idea of two-ness as an abstraction becomes a separate idea in the child's mind—

until this time the sound of the word 'two' is meaningless to the child, and only attains meaning when finally associated with this abstract idea of the two-ness of any two things. The meaning comes first; afterward the language symbol (word) is associated with the meaning and may be substituted for it when occasion demands. One has but to attempt to teach a child, who is just learning to talk, to count and deal with number concepts to see how absolutely meaningless the words are to the child until he has had opportunity in the course of his daily experiences to make the abstractions necessary to form these ideas. We may teach a three-year child to pronounce perfectly the sentence: "The square root of twenty-five is five," and if the action of language mechanisms constituted thought we should expect the child to understand perfectly the meaning of what he had said! Further comment seems unnecessary.

It will be noted that in this discussion it has been necessary to use concepts which are not found in Behaviorist psychology. These are the concepts of 'meaning,' 'idea,' 'concept,' 'conscious,' 'purposeful,' 'association of ideas,' 'abstraction,' 'symbolization,' etc. Yet these are fundamental to structural psychology and from the above discussion we deem it apparent that a consideration of the acquisition of language habits, their function in thinking, and the material of truly non-language thought, is totally inadequate without these concepts. To be sure we find passages in a Behaviorist psychology attempting to state what goes on in the mind of an individual. Thus (p. 305): "We manipulate vocally" (in attempting to recall the name of a familiar person) "by running over the names beginning with each succeeding letter of the alphabet, or by saying 'black hair,' 'blue eyes,' 'six feet tall,' and the like." This seems to the writer to be one of several excursions quite outside the realm of Behaviorist psychology. He does not know whence these ideas came but judges that it was by some sort of inference, partly because the Behaviorist does not use introspection and partly because he is unable to corroborate them by introspection.

7. *The Inadequacy of the Behaviorist Conception of Thought.* —To illustrate what is believed to be the wholly inadequate conception of thought as entertained by the Behaviorist, we cite the soliloquy postulated on page 332. "The implicit word processes (aroused by whatever previous stimulus) 'it's a fine day, I think I will go to the races; it's twelve o'clock now, I have just time to catch the train,' serve to start you to get your hat and field glasses. Some unfinished work meets your eye or other conflicting word processes are aroused, as 'but I have to write those letters and I have a luncheon engagement with X.' These tend to drive the organism as a whole into some other form of action; for a time there is a conflict (inhibition). Finally when the conflict is over the final word act issues, 'Well, I guess I'll have to give up the races and write those letters and keep my engagement with X.' Here we see implicit word processes tending to arouse overt acts and actually arousing the initial steps. But since the human individual is a completely integrated affair, associated word processes arise which may drive the organism into a totally different form of activity from that which was first initiated."

The writer contends that the 'previous stimulus' together with the mental activity which called forth this soliloquy would be sufficient to start one to get his hat and field glasses, without the accompaniment of any action of language mechanisms, and that such action itself would not suffice. The reasoning is as follows. Let us suppose the previous stimulus to be the perception of the green grass and sunshine and warmth of the outdoors. This perception called forth by association the memory of previous days when races were attended and of the accompanying pleasure. These memories contained the urge to renew the pleasures. They gave rise to the decision which is expressed in the language: "It's a fine day, I think I will go to the races." The decision made, the thought took form in vocal expression. At this point either the clock struck twelve, this serving as a stimulus, or the idea of going to the races naturally called up the idea of when to go, which in turn suggested the

idea of looking at the clock, resulting in the perception that it was just twelve o'clock. None of this mental activity required language. The perception of the time of day having been made in one way or another, the idea called up the words which would express it and the individual added, 'It's twelve o'clock now.' What happened next? Presumably at this point came the idea of going to the races by train, followed immediately by the idea which if expressed in words would be, 'When does the train go?' which in turn called forth the memory that the train goes (let us say) at twelve-fifteen. Immediately there came to the mind the idea of the preparation which is necessary to catch the train and a judgment is made as to how long this will take, based upon past experience. The individual must also go through a certain mental operation of determining how much time there is available before train time and make a comparison between these lengths of time in order to make the decision which when expressed in words is, 'I have just time to catch the train.' This idea possibly suggests the idea of haste which together with the idea of going to the races calls forth ideas of the appropriate preparation, getting the hat and field glasses, etc. These latter ideas take form in action.

In view of the obvious necessity for the mental activity of perception, judgment, decision, etc., intervening between the advents of the ideas which took form in the language quoted, we submit that, as stated above, it is impossible that the soliloquy postulated could of itself have given rise to the getting of the hat and field glasses. Moreover, the ideas themselves which suggested the soliloquy could have given rise to the acts and there need have been no language, explicit or implicit, involved whatever. Thus, the ideas of time may have been conceived in visual imagery—the imagery of the face of the clock and the movement or path of the minute hand. No language is required. The ideas of preparation for the train would consist of memory of the acts of getting hat and field glasses, walking or riding to the station, buying the ticket, etc., these consisting chiefly of visual and kinæsthetic images. No language is necessary.

The ideas involved in the judgment of distance (to the station) or of time required for preparation and traversing the distances, etc., would be kinæsthetic or visual or other ideas of space and motion, the comparison of these ideas of space and motion, etc., resulting in ideas of the relations between them. No language is required. Moreover, the ideas which take form in the acts of getting the hat and field glasses are visual, tactual, and kinæsthetic and are quite independent of language. The fact is, one could conceivably note the weather, decide to go to the races, make preparations, board the train, hand the conductor a ticket, note the progress of the train, get off at the race track, pay the entrance fee, and watch the races, all with mental activity and acts in no way involving language. Any amount of soliloquy or conversation may accompany the expedition, but this is wholly incidental, secondary, and unessential.

8. *Introspection.*—We believe that Behaviorist psychology is entirely sound *within its own sphere*, that is, so long as it confines its study to the behavior of the individual as seen from without. A psychology so limited, will, of course, necessarily leave untouched a vast field of useful knowledge which can in time be made scientific where not already so, after extensive investigation, comparison of findings, determination of general tendencies, and the careful observation of everyday experiences. But should one desire to explore the realms of psychology outside the scope of Behaviorism, he must then supplement his external observation by as thoroughgoing, extensive, and careful an examination of that which takes place within the mind—as seen from within—as is possible by highly practiced and trained introspection.

To direct the attention to the color of an object is a very easy matter. To direct the attention to the difference in shade between two colors may be slightly less easy but it is entirely possible. To direct the attention to the idea of the æsthetic value of the colors requires perhaps appreciably more practice, but it is none the less possible. However, to direct the attention to the nature of the mental process of choosing between two colors, and to the manner

in which the choice gives rise to appropriate acts, may be quite difficult, not to say impossible, for the inexperienced person. Yet these phenomena are available for observation no less truly than the habit of typewriting is open to acquisition or the length of a rail is capable of being measured to the thousandth of an inch. These accomplishments require long practice or minute observation, but we do not say they are impossible. An unpracticed person cannot direct his attention to the less tangible aspects of thought any more than he can play a theme on the piano. Because it is difficult however, one does not forego the learning of piano playing, if he desires to learn to play. Again, it is possible that no two observers might obtain the same measurement of a rail to the thousandth of an inch. Nevertheless we do not say that measurement is of no use in physics. Observers can nearly all agree on the length of a rail to the tenth of an inch, and on the length of a needle to the hundredth of an inch not to say to the thousandth.

Similarly in psychology, those inexperienced in introspection may not be able to distinguish between middle C on the piano and the C an octave above, or to observe that they see objects double which are not focused upon. Moreover, persons highly trained in introspection may not always be able to distinguish between the perception of a very faint sound (as the distant ticking of a watch) from the auditory image of the sound (imagined sound) nor to state just what constitutes the mental element of difference between the emotions of fear and anger. But there should be little difference of opinion between persons of extended experience in introspection as to whether the material of our thoughts, when we create new ideas, and conceive new modes of activity in the fields of music, art, drama, mechanics, etc., is in the form of language or in the form of tones, visual pictures, etc.

9. *Summary.*—There may be no experimental proof whether or not thinking—conscious adjustment to the environment—is invariably accompanied by the actuation of some language mechanism as the larynx, lips, fingers, etc., in the incipient production of some form of language, spoken

or written, but the evidence would seem to favor the belief that no such invariable accompaniment is necessary. One uses the eyes in observing objects attended to almost throughout the waking state. It would seem more plausible to assert therefore that some form of "implicit" *eye* movement is a necessary accompaniment of all thinking. Much evidence such as that from the observation of a chess player studying his moves could be brought forth in support of this view. All this, however, is quite beside the point. The claim is made by the Behaviorist that "thought *is* the action of language mechanisms" (italics mine). Certainly the evidence against such an assertion is overwhelming.

Man is an organism highly adapted physiologically to his environment, provided with sense organs of sight, hearing, taste, smell, touch, pain, heat, cold, muscle movement, body position, etc. Each of these sense organs is capable of giving rise to sensations which take the form, in the mind, of images or image patterns. (The word image is used in a very broad sense as shown below.) The organism has at its disposal any or all of these incoming percepts or stored images or image patterns as material for thought, for working over into new combinations, new thoughts, which will give rise to new actions, new adaptations to the environment. In the event of the bringing together of two or more concepts—images—or of the dividing of one concept into two or more (as when a child first separates from the concept ball the concept roundness)—in the event of this working over of concepts, they are necessarily abbreviated, composited, exemplified, attenuated, or substituted for by others. If the substituted concepts are of a kind remote from the kind for which they are substituted but are more or less definite and commonly understood, we call them symbols. A careful description of the manner in which thought material is abbreviated, composited, attenuated (even to a point which is considered by some psychologists to be 'imageless'), etc., is of course quite impossible within the limits of this article.

However, the mental activity which brings forth a new act may be the result of the combination or division or other

working over of any type of mental material—the bare sensation, the fresh vivid full percept, the fairly vivid memory image, or the image or image pattern when abbreviated, or attenuated, or composited, or exemplified, or in any manner generalized or particularized, or finally in the form of symbols. And language, as has been shown, is but one general type of symbol system.

In conclusion, then, let it be said that we *may* think in words, and when we do, the thinking may be accompanied by the action of language mechanisms. But thought—even conscious mental adjustment—is not restricted to the material of language any more than it is restricted to the material of musical tones or of architectural designs or of facial expressions, nor is it restricted to the action of language mechanisms any more than it is to the mechanism of hearing or of sight or of locomotion.

A BEHAVIORISTIC ACCOUNT OF SLEEP

BY CHARLES H. WOOLBERT

Division of Public Speaking, University of Illinois

The phenomenon of sleep does not lend itself conveniently to explanation in terms of sensation, image, and feeling. Accordingly structural psychology can offer little by way of a description of the state of sleep. The behaviorist, on the other hand, believing that his definition of consciousness offers a statement not only of what consciousness is, but of what it is not, is in a position to explain sleep. If, as behaviorism asserts, mind is a matter of reflex connections always involving the movement, tension, or tonicity of muscles and always correlated with the activity of glands, then consciousness, in its varying degrees of clearness, is a matter of degrees of complexity and ordination among systems of muscular action. A high degree of consciousness thus becomes synonymous with an intricate and ordered complexity of tonicity in muscular systems; while a low degree of consciousness is equivalent to a complexity of low degree and an ordering of simple texture. Consciousness thus may be, speaking in a paradox, scattered, involving perchance abundant activity—by way of muscular tonus in various muscle systems—but activity of a low degree of intricacy in organization.

If consciousness, then, be a matter of the degree of complexity of interacting muscle systems, non-consciousness is a lack of activity or else a lack of this complexity. In either case the factor of complexity is vital and needs describing. Behaviorism's explanation of this concept is based on the continuative function of the sense endings within the muscles. Always stimulated by any muscular event, they afford the means for causing a single inaugurating stimulation to reverberate through a long series of tensions, or else to provide a continued hardening of some one set of muscles. The type of tensions running in a series is what is called the chain

reflex; the continued is that called circular. Chain reflexes have much to do with that complexity of interacting muscle systems that makes consciousness. Carried on by means of proprioceptor organs of stimulation, they have a right of way and a special kind of clearance. This insures them continuity and lends a measurable degree of stability to the complexity of structure that makes up consciousness.

This description recognizes, first, that the muscularity of the body occurs in fairly well-defined systems: as those of the back, the legs, the head, the face, and the throat; and, secondly, that these muscle systems are set off one by the other and in a certain order. This order is based on priority. Priority of muscle systems is, in general, a matter of precedence in the development of working efficiency in reflex arcs. Those systems that have an early development history come to have a pronounced control over systems developed later, in that the systems developed later get their initial determinations from the workings of the habitual responses of the earlier. Their most intense determinations are, both earlier and later, then, conditioned by the determinations of the systems already habituated. Their capacity for quick and valuable response depends thus very largely upon their close coördination and coöperation with habitual reactions of systems determined at an earlier stage of development.

An understanding of sleep requires a description of this development order, a description of gross muscle systems and a statement of their superordination and subordination. This is found in an account of the operation of Pawlow's Law, as manifested in the conditioned, chain, and circular reflexes.

By Pawlow's Law a reflex arc may be a factor in new activities by the stimulation of sense endings imbedded in the muscles that are contracted by the operation of such an arc. Every motor process stirs a muscle; this stirring starts new impulses; and these impulses seek a new outlet. Two directions they can take: they can go around back by way of the motor nerve that stimulates this muscle, and so stimulate it again; or they can take some other motor nerve leading to a

quite different muscle. The former of these processes is the circular reflex, the latter, when carried on through a series, the chain reflex. By means of circular and chain reflexes in various combinations the organism has within itself the machinery for carrying on activity, for a while, at least, without the intervention of peripheral stimulations. Only in this way can the continued activity of involuntary organic acts be accounted for, activities like heart-beat and breathing. In this way also is given a satisfactory account of the mechanism of such activities as catatonia, catalepsy, emotional complexes, fixed ideation—in fact, repetitive and continuative actions of all kinds.

To get the full significance of these circular and chain reflex systems in the coördination and superordination of actions, it is necessary to envisage them in connection with the development order of muscular systems. Systems developed early in the life of the organism are obviously determined strongly; in the case of the very earliest, heart-beat and breathing, no stimulus short of that adequate to stop life can divert them. Moreover, their continued activity is but little dependent upon other stimulation than that provided by the circular reflexes that keep them regular; at least so long as the muscles concerned receive nourishment from the blood. Systems developed one stage later, like those involving walking, reaching, turning the head, have within themselves much of the same continuative mechanism. Their ability to function, however, they have gained largely as an adjustment from the successful functioning of systems developed earlier. At the start of their functioning they do not possess a full measure of self-determination; they are of necessity dependencies, subject to the caprice of superiors holding power by a rule of seniority. Certain things they can do so long as the older systems go about their business in an orderly fashion; certain other things, under the same conditions, they are not privileged to attempt. Thus the use of the eyes, the hands, the legs, is conditioned very materially by the regularity of the beating of the heart and of breathing. Let, once, something go wrong with either of

these older activities, and the organism loses complete control, eventually, of hands, legs, eyes, and of all other muscle systems. Under similar circumstances any determinations that are ordinarily well established give way to a recrudescence of the wildest of random movement.

Thus the muscle systems operate in a kind of hierarchy, with jurisdictions fairly distinct, though not exclusive. Most firmly enthroned of all are the primary reflex systems controlling heart-beat, flow of blood, operation of vital organs, and breathing. Next come those developed in the organism's earlier days, use of arms and legs, back, torso, and neck muscles; later, and probably overlapping the earlier systems, muscles of the eyes, ears, face, and head; lastly—coincident with the development of speech—the muscles of jaws, lips, tongue, and throat. Thus consciousness as complexity of muscle systems is a pyramid with the organic systems at the base and the muscles of thinking, reasoning, and speech at the top. Or, changing the figure, it is a hierarchy with the organic systems as autocrats and the other systems holding office on a descending scale of self-government, dependent always upon the commands of the autocrats ruling by virtue of prior possession of power.

This hierarchy operates to provide the difference between sleep and waking consciousness. Without the tension of head and face systems there is not complexity enough for consciousness. So vital are they to clear cognition that they are easily confused with the totality of consciousness; remove them altogether from the systems active at any one time, and unconsciousness occurs. Yet they are not autonomous; when fatigued and free from intense peripheral stimulation they normally yield easily to the relaxing of the lower systems and go out of function along with them. Any condition in which they refuse to stop functioning when free from peripheral stimulation or when fatigued, and when the lower systems have relaxed, is looked upon as abnormal. In fact psychopathic conditions can be described either in terms of an actual lack of the upper systems, or in terms of their failure to coöperate with the activities of the lower. Sleep is accounted

for in the formula: Remove the higher systems from activity, and consciousness departs altogether; weaken the lower, and consciousness is in a precarious condition, especially if the higher systems are affected by fatigue. When the lower systems are thrown out of function, the higher circular reflexes either stop at once or, in abnormal cases, ultimately wear themselves out; in either case consciousness breaks up. Remove the lower entirely, and death is instantaneous.

A prime requisite of easy and deep sleep is freedom from stimulation for the eye, ear, nose, tongue, and parts of the skin not constantly pressed by clothes; forms of stimulation that have little to do with the organic systems. Yet sleep is possible even in the face of such stimulations; but only in cases where great fatigue throws lower systems, like those of leg muscles, back, and neck, out of commission. The chief power of estopping other systems, especially under conditions when fatigue is present, is authoritatively appointed to the organic systems; because they get their determination at a time when the organism is in its most plastic state—in its early stages.

Thus sleep becomes behavioristically a matter of the efficient domination of the upper systems by the lower, operating through the relaxing power of fatigue; while wakefulness and insomnia always imply that the higher and later systems are assuming dominance over their precursors. Wakefulness, so, is characteristically the dominance of the lower systems by the upper when fatigue is not present. Accordingly when wakefulness exists at the same time that fatigue is present, the condition is abnormal.

This means that when the muscles of the back, legs, and neck are relaxed, a powerful stimulator is lost to the muscles of the arms, hands, feet, and head. When arms, hands, feet, and head muscles in turn are relaxed, there is lost a powerful source of stimulation to the muscles of the face, jaw, tongue, and throat. While systems are undoubtedly more finely differentiated than this, still their hierarchical interdependence is on just such an order—and sleep can be explained by gross characterizations as well as by those more minute. In the

inducing of sleep much significance must be attached to the order in which muscle systems go out of function. When the organism follows the development sequence, sleep is easy; when the order of relaxation is in any way reversed, restlessness and wakefulness follow. In cases of complete reversal of the order, we get such states as hypnosis, temporary high degress of attention, manic conditions, and forms of insanity.

The beginning of sleep then normally is the relaxation of the muscles that hold the body erect. As soon as these muscles are relaxed, the prime determiner of higher systems is taken away, the proprio-ceptor foundation, and the higher systems then are kept in function by only a veritable bombardment from the outside world or from very strongly determined circular or chain reflex arcs within their own system. Among these latter are emotional states, fixed ideas, tunes running through the head, repeated attempts to solve a problem, rhythmical verbalizing, and thinking in circles.

The next step in normal sleep is the sequential relaxation of each of the systems hierarchically dependent upon the erect-holding systems. Finally through the sufficient dissolution of the complexity that makes consciousness, sleep comes. Consequently once a person lies down, relaxing the muscles of legs, back, and neck, the beginning is made of sleep. Providing there is no interference from outside stimulations—chiefly those acting upon the sense endings in the head—and also so long as there is no intense circular reflex process going on in the muscles of the jaw, lips, tongue, or throat,—'thoughts that will not go out of one's head'— such a beginning once made leads to complete sleep and loss of consciousness.

Certain easily-made empirical observations as to sleep confirm this account. (1) Sleep is characteristically accompanied by relaxation of muscles. (2) Characteristically also it takes place with the body in a horizontal position, a position that induces first of all a relaxation of the muscles of the legs, back, and neck, all of which must maintain a high degree of tonicity to maintain erect posture. (3) When the muscles of the back, legs, and neck are relaxed, the muscles of the

head, face, jaw, and throat all tend to relax in a short time ensuing. (4) These muscles also are most easily relaxed when freed from stimulation of the head sense endings, in the dark and in silence and free from intense taste or smell. (5) All these relaxations occur parallel with a scattering of consciousness, a defocalizing of attention; and the greater the degree of relaxation, the less the ability to perform any act implying a high degree of concentration. (6) Organic activities, though, are kept up despite any relaxation of the voluntary muscle systems. (7) Again, deep sleep implies complete relaxation; also it connotes rest and recuperation from fatigue, a retoning of muscles for future work. (8) The degree of sleep involved conditions the number and vividness of dreams had; deep sleep implying few dreams and light; light sleep implying many dreams or dreams that are vivid. (9) Deep sleep also leaves little recollection of dreams, except for the moment when one is coming out of sleep to consciousness. (10) Great numbers of dreams, or dreams that are vivid, it is generally assumed, are equivalent to defective rest, and restlessness is always accompanied by the inability to stop thinking or by numerous and intense dreams. (11) During widespread muscular activity there is no such thing as sleep; as during walking, eating, reading, talking, giving active attention in any way.

These general observations, and many others of similar nature, point clearly to the close relation existing between sleep and the movement or tonicity of muscles.

In the subjective terms of ideation, sensation, and feeling it is difficult to explain what happens to conceptual thought during sleep. The behaviorist, by assuming that thought of all kinds and in all degrees is a matter of muscular tonus, precisely as in walking or standing erect or moving the hands or talking, can give an account of sleep that fits in with his whole program. Sleep to him is nothing but a disorganization of muscle systems which in waking consciousness are closely interrelated hierarchically, the action of each determined in part by the continued action of the others. When conditions are set for the relaxation of sundry systems of

muscles, consciousness begins to be more scattered, system after system drops out of function, and ultimately, in the soundest sleep, nothing is left by way of muscular activity but the functioning of the organic systems.

Dreams are clearly the result of systems involving throat, face, tongue, and lip muscles which remain in function when other systems have been thrown out of gear, systems which, if combined, would make consciousness. The Freudian dream psychology presents agreement with obvious facts in that it recognizes the existence during sleep of mental processes which seem very like others that go on in waking consciousness, yet which at the same time are partly unlike them. The behavioristic explanation of this is that in so far as a dream is a matter of the activity of muscle systems involving a high degree of complexity and coördination, in so far it is similar to waking consciousness; and so has a way of seeming logically coherent. The illogical dream, on the other hand, the freakish dream, the dream that seems to forbid explanation and interpretation, can be explained broadly as a type of organization and coördination not met with in that form in waking life; so that, speaking generally, the more unusual the combination left operative during sleep, the more fantastic the dream. From these suppositions can also be found the reason why dream analysis cannot be a matter of accurate interpretation and why the Freudians who assume to interpret all dreams give promises in reality beyond powers of observation to fulfill. So entirely beyond inspection and prediction can be the permutations and combinations of the hierarchy of muscle systems, that they can defy all powers of analysis.

From these observations follow certain therapeutic inferences worthy of note, most of them current already through the experience of the race. If you would sleep soundly, exercise much, in particular the muscle systems of the body below the head; for if the 'body' is tired, the 'mind' will rest also. If restless in sleep, study how to relax, first of all the muscles of the legs, back, and neck. Then reduce the breathing rate: high tension almost invariably is accompanied by rapid breath-

ing; low tension by slow-breathing. Also hands and feet, fingers and toes, must be inert. If thoughts crowd thick and fast and will not leave, let the jaw drop, make the muscles of the cheeks and lips flabby, avoid screwing up the muscles around the forehead and the eyes, see that the tongue lies limp in the mouth, and make certain that the muscles of the throat are not in any way tensed. These last-named muscles, together with those of the jaw, tongue, and lips, are more likely than any other to get in the way of sound sleep. Make sure to observe the right order of relaxation of systems; gross lower systems first, then the finer systems below the head, and finally the fine systems of the head. Sleep is synonymous with carrying out the following order in relaxation: Reduce the breathing rate; then relax legs, back, abdomen, and neck; then arms, hands, fingers, and toes; next the muscles around the eyes, forehead, scalp, and ears; and finally those around the mouth, jaws, tongue, and throat—the muscles of speech and conceptual thought.

THE COMPENSATORY FUNCTION OF MAKE-BELIEVE PLAY

BY EDWARD S. ROBINSON

The University of Chicago

It is the purpose of this paper to present in outline a view of play which will usefully supplement those theories which are generally entertained. Play is, of course, a phenomenon of extreme complexity and, for that reason, its complete explanation should hardly be looked for in any single statement. The suggestions have been made that the play of children is a chronicle of race activities, that it gives practice or preparation in functions of coming importance, that it furnishes an outlet for surplus energy, and that it is a recreational agency or means of relief from fatigue induced by other occupations.[1] Undoubtedly all of these things are true to some extent, but more important than any of them is the fact that play is essentially a compensatory mechanism having the same origin and impetus as the day-dream or fantasy.[2]

A compensatory function is especially evident in that type of play which involves the element of conscious shamming or make-believe. While it is possible to demonstrate that other types of play may operate as compensation,[3] I shall confine the following discussion to play which is clearly make-believe in character.

The child is driven by many inherited and acquired impulses, some of which are adequately and easily expressed and some of which find no direct outlet. These latter create a situation demanding compensation, and this compensation

[1] For one of the latest discussions of the various theories of play see Reaney, M. J., 'The Psychology of the Organized Group Game,' *Brit. J. of Psychol.*, Monog. Supplement, 1916, IV.

[2] The compensatory nature of the day-dream or fantasy is clearly brought out by Freud and other writers of the psychoanalytic school.

[3] Dr. Reaney, for example, holds that organized group games may have a compensatory function. *Op. cit.*

is as a rule secured through make-believe activities. Most common among such activities are play and fantasy. A child would fight, hunt, and make a home as particular stimuli arouse him. He is seldom in such an environment, however, and he is practically never so organized by inheritance or training that these undertakings can be fully carried out. There are inexhaustible inhibitors around him and within him which check free expression. And so he plays at, or has day-dreams of, fighting, hunting, and home-making. I have no desire at this time to say which of the unsatisfied impulses of childhood are inherited and which acquired; but, however they arise, we find that they are many and urgent, and consequently that every normal child must find compensation for their inhibition.

There are a number of factors which may act as inhibitors of the behavior tendencies of children. These may conveniently be divided into the *extra-organic* and the *intra-organic*, according to whether they are in the nature of environmental interferences or interferences which arise out of the child's own organism.

During his development the child is constantly running into *extra-organic* or environmental facts which are incompatible with the satisfaction of his desires. He may want to hunt. Perhaps the family cat supplies him with a stimulus to make this impulse felt. But this hunting impulse has become a particularized affair. Hunting is shooting, and he can not shoot because he has no gun. Instead of ignoring a stimulus to which he can not react adequately, he points a stick at the cat and shouts 'Boom!' He may then, and perhaps to his sorrow, try to drag in his 'dead' game by the hind legs. But the main and incontestable point is that the child is compensating, by means of his pretensions, for the inadequacies of the situation. He would like only too well to shoot a real gun and drag in game which is really dead, but his environment does not supply the appropriate circumstances. And so he plays.

Among the more important *extra-organic* factors which limit the child's expression are the people around him. Just

as he discovers the splendid interior of his father's watch, someone takes the watch away from him. Just as he discovers the importance of certain corners of the pantry, someone carries him away to another room. Everywhere there are people and they are constantly interfering with his behavior.

As I have intimated, it is not only the lack of a physical world fitting in with every whim which causes the child to play rather than to act in earnest. He has also his *intra-organic* interferences arising out of his own complex little nature. For the pure joy of it he would, at times, like to bring down a stout club upon the head of his playmate—that is, he would like to do this if it were not for the disconcerting facts that he would not like to hear his playmate cry in pain, and that he would not like to feel the blows of his playmate's revenge. And so the two boys will play at fighting. Often, too, a child is hindered from acting as he would because of a realization of the smallness of his body and the slightness of his muscular strength. In such cases we are apt to have a mimicry of feats of strength and daring.

It is evident that there are instances of make-believe play and fantasy which apparently, at least, are not primarily compensatory. A child may straddle his hobby-horse, not because it is the best substitute for a real horse he would ride, but simply because he has been taught to do so by his parents. There is little doubt, however, but that the average child enjoys his playing the more where he perceives its symbolic relationship to a more serious pursuit. The fact that children's play is given much of its specific form by adults, does not, in the last analysis, indicate that it is therefore less compensatory. By custom and tradition we initiate various make-believe performances for children, but something in the nature of childhood must explain why children take to the make-believe with such enthusiasm. When we first teach a child to ride a hobby-horse he may be unaware of any connection between this activity and the actualities of horseback riding. But as he learns about real horses and real riding, his play will become more and more clearly com-

pensatory in function. In other words, the rise of certain impulses in children is so inevitable that their compensatory expression may be provided for by the customs of the race. In the case of any one child a compensatory activity may be set up before the need for that particular compensation arises, but we may still consider the activity a typical product of child life and its characteristically incomplete adjustment.

Just as in certain individual cases a compensatory make-believe may arise before the need for that particular compensation, so specific habits of play and fantasy may be retained after the apparent need for their compensatory service is past. I know of successful men who find great pleasure in day-dreams of achievements which they would not care to have realized in any tangible fashion. In some of these cases the day-dreams express real desires which are denied direct expression because they run counter to other desires of a more powerful sort. In other cases, however, it is quite possible that day dreams which once had a compensatory function now operate as old habits and are retained because of their own repetition rather than because of any important compensation which they still render.

Play and fantasy are frequently concerned with situations more painful and disagreeable than any we should choose to meet in real life. A natural question arises as to the sense in which such make-believe can be considered compensatory. Children do not want to be in railroad wrecks nor to receive bullet wounds, and yet they enjoy pretending they are in such straits. So it appears on first thought, but, as a matter of fact, children do wish that just such things would happen to them, providing they might happen without pain or other ill consequences. In regard to railroad wrecks, if we could read a child's impulsive nature completely, we should probably find that he wishes he could be in a wreck and, at the same time, hopes he will not. He is in the same predicament as the boy who would like to club his companion and yet would not like to. And like that boy he compensates for his conflict by playing. In other words there are few, if any, situations in life which appeal to us in a purely negative way.

We do not, as a rule, want to suffer great misfortunes; yet there are certain factors, such as affectionate demonstrations on the part of our friends, the joy of being in the public eye, and the like, which give the majority of unfortunate circumstances a considerable amount of positive appeal.

Distinctly unpleasant play and fantasy may also provide for the compensatory expression of negative impulses. There is little reason to believe that fears, for example, do not require expression of some sort as urgently as more positive tendencies. Playing, day-dreaming and the telling of stories involving ghosts and goblins may well serve to express fears which must be inhibited in the world of actuality.

Holding the older view that childhood is a period of happiness and serenity, one could hardly accept an explanation of play in terms of compensation for incomplete or faulty adjustment—in terms of the partial resolution of conflicts between the child and his environment or between contradictory factors within his own character. I believe, however, that there is little need to argue against that older view. Childhood is primarily a period of incomplete adjustment, and we remember it as peaceful because we have forgotten its sorrows and because problems of great consequence to us in childhood mean little to us now. Full of impulses to do actual things, the child is equipped with a physique and surrounded by an environment which are constant obstacles. I do not believe, like some, that it is desirable, if possible, to remove these obstacles and make childhood a comparatively easy and comfortable state. Human life requires, and gets much of its value from, an abundance of nice adjustments which can come only as the result of long and necessarily arduous training. The child comes into the world with an inherited behavior equipment, but at best this equipment is an uncertain affair. Each impulse tends to operate in inappropriate as well as in appropriate situations. Each impulse, if the child is to become prepared for adult life, must be defined, and definition implies inhibition. The child must live through a period of paradoxes before he can become an individual of discrimination. If he were a perfect mechanism

and if educational stimuli were perfectly coördinated, it is possible that he might be trained without being constantly thrown out of adjustment. Then, too, if the life for which society prepares him were more simple in its requirements, he might be spared some maladjustment. But the human organism is not perfect, and, while educational practice improves from time to time, the world rushes forward into new complexities. One who has any faith in the present direction of progress can hardly do other than accept the essentially incomplete adjustment of the young as a necessary product of that progress. The happy fact is that the conflicts of youth can be so adequately compensated for by the play and fantasy mechanisms.

While they are fundamentally natural and necessary phenomena of child life, play and fantasy can result in pathological as well as in normal compensations. And as normal compensations the forms which they take may modify the development of character to a marked degree. For these reasons their exhaustive study, as but different manifestations of a single process, is essential. Indeed it seems to me that few fields may be more profitably explored either by those interested in child life in general or by those interested in some particular child. From this compensation process, studied for what it really is, we may hope to gain some new and useful knowledge about the stresses and strains of human development.

Still, even at the present time, it is possible to point out some of the principles which operate in the compensatory behavior of children.

The distinction between play and fantasy is, of course, a distinction between overt and ideational behavior. Play, in so far as it is pretending, is never without an element of fantasy, but we may arbitrarily confine the application of the latter term to those forms of pretending which are lacking in overt bodily accompaniments. It will then be possible to distinguish between these two types of compensation and to note their interrelations.

Although we cannot be certain of it, play probably pre-

cedes fantasy in the child's life. The latter does appear quite early, however, in some children at least, and before the school age is reached both are clearly present. I remember the interesting evidence for the early rise of fantasy given to me by a little girl of not more than four, who said, placing her two chubby hands before her face, "Let's shut our eyes and play we're at Gran'ma's."

In play and fantasy there are two factors, which may or may not be consciously recognized by the child, determining to a large extent the nature of his pretending activities. In the first place, there is a tendency toward breadth and freedom of expression. The child must express impulses which are often clearly incongruous with his world of actuality, and the greater this incongruity the more lively will be the flights of imagination to which they give rise and the more apt will the child be to engage in private fantasy rather than in overt play. In the second place, the satisfaction which is derived from compensatory behavior depends to some extent upon its being within the limits of the child's own credulity. The impulses which drive the child are aimed at an actual world, and their indirect expression itself must not get too far beyond the realms of that actuality. Thus, we may think of these two main determinants of play and fantasy as (1) the child's natural tendency toward free expression, and (2) his need for a certain credibility in experience.

The tendency toward free expression leads to the establishment of all sorts of fictitious characters and objects within the playground. Toys and playmates which do not fit in with the completer, fancied world may be put aside. I remember that even up to the age of sixteen I frequently judged congeniality in terms of the readiness of others to disregard reality in favor of a world of pretty definite and well defined fancy. I always preferred to *knock grounders* with one particular lad because he coöperated so well in converting the procedure into the pretensions of a *big league* game. The same was true in boxing. Having read and memorized the details of most of the historic ring battles, we repeated many of these almost blow for blow upon the floor

of my mother's laundry. And many were the Harvard-Yale football games in which I engaged with one other actual player, both of us, as often as not, playing on the same side. In cases of this sort, the meaning of ordinary play activity is widened by the liberal use of fantasy.

In the course of an individual's development many impulses arise which can not be expressed to any satisfactory extent in a coöperative fashion. Often a child is afraid of being laughed at for the world he would live in. Under such circumstances there may be a withdrawal from play to pure fantasy with its wider possibilities for pretending. Indeed, one of the signs of coming adulthood is the giving up of overt play and the switching over to compensatory behavior of a more private sort. Adults seldom play in the childhood sense of that term, unless it be in art. In the adult, compensations through pretending are more likely to be worked out in private day-dreams. The fact remains, however, that less compensation of any kind is necessary in the general run of adult lives, so that we may safely assume that fantasy as well as play is more common during childhood.

Along with this tendency toward free expression, we have a tendency to make that expression as realistic as possible. Children are constantly recognizing inconsistencies in their play life and trying to patch them over as best they can. When, as a very small boy, I played with tin soldiers and miniature locomotives, I always felt the inappropriateness of the size of my own body. The device which I hit upon to get around this difficulty I called *Playing You Are Nothing.* Every playfellow who entered into the world of my tiny armies and railroads was introduced to the proposition of suspending all interest in his own body. The running of the trains and the marching of the troops were to be considered as events independent of ourselves. There was one youngster who could not push a locomotive across the floor without playing *he* was the engineer. His fate was obvious. I never invited him to play unless I could get no one else; and, when he did come, it was to be made miserable by my constant insistence that he must play he was nothing. Our dis-

agreement, of course, grew out of the fact that each of us in his own way was striving to give the play a more vivid atmosphere of reality.

Just as overt play often passes over into private fantasy owing to a struggle against the limitations of the actual social and physical world, so private fantasy often passes over into overt play in the interests of greater credibility. As a child I was full of baseball fantasies. Although I played baseball a great deal, these games did not satisfy certain standards set up by reading athletic stories and watching older and more skillful players. But the fantasies, too, often became unsatisfactory on account of their intangibility. As a result I formed the habit of laying out a diamond upon the lawn and there, without ball or playmates, carrying out the overt movements of an heroic baseball performance. Many a time, I pitched nine long innings to baffled athletes who swung immaterial bats at my imaginary curves. Here was fantasy improved and made realistic by the actuality of its muscular accompaniments.

The topics of private fantasy are perhaps even more apt to find increased tangibility by being brought into contact with a real social world. The child knows that his day-dreams are unreal, but the insistence of that fact becomes less troublesome if only he can get some one else to believe or act as though he believes in the reality of those imagined events. Many of the lies of children arise out of such circumstances. A boy longs for a pony and a box of tools. He fancies these things in his possession, and before a great while he somehow feels driven to tell his friends either that he already has the things he desires or that he has been promised them. An acquaintance of mine spent her earliest years on a farm which was more or less out of touch with the livelier affairs of the world. Now it so happened that an older sister in this household was sent to town to finish her education. Upon her return she had much to say of her experiences. These tales thrilled the younger sister and stimulated her to day-dreaming. Soon after this the little girl began her own education at a neighboring country school. As she tells of

it now, almost her first intercourse with her school mates was marked by her own spectacular reports of what *she* had seen and heard while sojourning in the town which really she had never been near.

It is interesting to note here that the literary make-believe of adults contains within it evidence of the tendencies toward free expression and credibility, which I have mentioned as such significant factors in child life. Written fiction, for example, may be thought of as an instrument for free expression and the spoken drama as an instrument for giving human fancies increased tangibility. It is hardly necessary to point out the importance of artistic appreciation and production for the compensatory life of children.

This view of play as a compensatory mechanism does not pretend to refute the more familiar theories which, by the way, were not formulated with special reference to the make-believe. That theory which describes play as a recapitulation phenomenon simply states that the primary impulses expressed in play appear more or less spontaneously at set periods in the child's life, and that the child's activities during successive periods of his life are definitely reminiscent of the typical periods of racial development. Most of us would probably admit that there is a rough similarity between individual and racial development. But the view that play is a compensatory activity demands neither the acceptance nor the rejection of this theory. One need not know nor try to guess the exact origin and analogies of a child's impulses to realize their variety and the conflicts among them which demand the compensatory service of the make-believe.

The theory that play prepares a child for later life, if broadly enough interpreted, fits in quite well with the notion which has been developed here. Many impulses arise during childhood which, while they can not be directly expressed at that time, still demand preservation. A boy may be interested in machinery. If he is permitted or even encouraged in his play and fantasy to concern himself with machines, a very useful interest may be preserved for the time when it can find adequate expression. If it were not for compensatory expression through play and fantasy, it is quite conceivable

that many such early rising interests or impulses would suffer repression and thus be lost as far as useful functioning is concerned.

The theory that play furnishes an outlet for surplus energy is somewhat vague, but as far as it goes it meets with no contradiction from the conception that play is compensatory in function.

Much the same may be said for the recreational theory, which really finds some little support in the type of facts which I have been presenting. A boredom which longs for some impossible or impractical distraction is often indirectly relieved by a compensatory make-believe. The school boy, tired of his lessons but afraid to dash from the class room, may partially satisfy himself with a fantasy of the swimming hole. The worried business man, whose unused muscles would not tolerate exertion, may yearn to play ball and take his yearning out in fantasy.

In conclusion, play, the more private forms of fantasy, much lying and story telling, and the appreciation of stories all serve the same fundamental purpose in human life. They are compensatory mechanisms. They are more typical of children than of adults, because it is in children that the most incongruity exists between different impulses and between impulses and the surrounding world of actuality. The nature of play and the other compensatory mechanisms is determined by the need of imperfectly adjusted organisms to express their impulses as freely as possible without too greatly straining the possibilities of their own belief.

It is essential, if not self-evident, that play should not be thought of as behavior which is usually undesirable or pathological simply because its function is compensatory. Neither should we think that, because play grows out of imperfect adjustment, we should strive for a world in which play is unnecessary. Simpler organisms than ourselves get compensation through play. The ancients in a comparatively simple civilization got compensation through play. And in all likelihood the further humanity advances upon its present path of progress, the more important will be play and its related phenomena, especially for the young of the species.

THE CONTROL OF ATTITUDE IN PSYCHOPHYSICAL EXPERIMENTS

BY EDWIN G. BORING

Clark University

Dr. Godfrey Thomson has recently published some very illuminating discussions of the mathematical logic of psychophysics, and I for one feel myself too much in his debt to to indulge in the mere picking of flaws. Nevertheless his paper, 'A New Point of View in the Interpretation of Threshold Measurements in Psychophysics,'[1] gives me concern because it seems to cast aspersions upon what I have regarded as the most promising direction of development in psychophysics. It is true that Dr. Thomson's 'offences' are implied rather than actual and that ultimately I may discover him in agreement with my thesis; nevertheless his article furnishes a reason for placing the point of view that I have in mind squarely before psychologists. I may add that I conceive that I am merely explicating an idea that arose within the Cornell Laboratory. If there be any credit for its origin it is due Cornell, though the responsibility of the present exposition is mine.

Dr. Thomson's suggestion is as follows: In the test of Weber's law (he is thinking of lifted weights as the example) we may take the differential threshold, which is half the distance between the upper and lower limens (*i.e.*, half the 'interval of uncertainty'), as the measure of sensitivity; and such has been the usual practice. Dr. Thomson, however, prefers to take as a measure of sensitivity the interquartile range of the point of subjective equality. The threshold is proportional to the distance between the two points where the psychometric functions for 'greater' and for 'less' cross the 50 per cent. abscissa, and its amount

[1] Godfrey H. Thomson, 'A New Point of View in the Interpretation of Threshold Measurements in Psychophysics,' PSYCHOL. REV., 1920, 27, 300–307.

depends upon the number of judgments that fall within the third category of 'undecided' judgments (as Dr. Thomson styles them) or 'equal' judgments. Dr. Thomson would take these 'undecided' judgments and divide them equally for every stimulus value between the 'greater' and 'less' categories, thus establishing two new psychometric functions (instead of the original three) which of necessity will intersect upon the 50 per cent. abscissa and give zero limens. The interquartile range is the distance between the two points of intersection of these new psychometric functions with the 75 per cent. abscissa. It is independent of the 'undecided' judgments and dependent upon the measure of precision (h) of the psychometric functions, since the steeper the curves the less the interquartile range, and *vice versa.*

Dr. Thomson's preference for the interquartile range as a measure of sensitivity lies in his distrust of the relative frequencies of the 'undecided' judgments upon which the threshold depends. The threshold, he writes, 'depends entirely upon the subject's readiness to give the answer *undecided.* It measures therefore rather a moral character than a physical sensitivity.' 'The moral character of the measure $S - S'$ is above all seen from the fact that any subject who wishes may reduce it to zero, whatever may be his actual sensitivity, simply by determining that he will never give an answer *undecided.*' Thus the interquartile range 'is more physiological than the threshold measure.' The threshold is to be mistrusted because 'the decision as to what proportion . . . is to be called heavier, what undecided and what lighter depends upon a conscious act of the subject, and can be varied, if he be so disposed, at his whim; and will vary with his mood at the moment.' It is such a designation of things constant and measurable as 'physiological' and of things inconstant and uncontrolled as 'moral' or psychological that moves a psychologist to reply.

It is not my present purpose to inquire how much of Dr. Thomson's argument is actually new. Fechner wished to measure *Unterschiedsempfindlichkeit* by h and divided his

zweifelhafte cases between the 'greater' and 'less' categories.[1] G. E. Müller opposed Fechner, arguing that the limen must be used as the basis of Weber's law.[2] But Fechner stuck to his guns,[3] as did also Müller.[4] In 1904 Müller was wondering 'wie in aller Welt kann man ohne weiteres voraussetzen' that the measure of precision, which is independent of many factors entering into the lifting of weights, could constitute the basis of Weber's law. Dr. Thomson may need to meet Müller's argument against Fechner.[5] I belong to the younger generation to whom Weber's law and *Unterschiedsempfindlichkeit* are less sacred than they once were, and I am willing to admit that the matter may well wait for supporting facts. Experimental studies are wanted that show both the threshold and the interquartile range as functions of the absolute magnitude of stimulus, and then we can determine how each fits the Weber-Fechner formula. But I am not willing to let Dr. Thomson dissuade us in advance from an interest in the threshold because its 'morality' can not be controlled. There is enough experimental work to indicate, so it seems to me, that accurate control of the third psychophysical category is possible and scientifically necessary.

The Experimental Control of the Psychophysical Judgments

1. In the first place, if we are to gain accuracy of definition of the psychophysical categories, we must exclude the doubtful judgments.

Dr. Thomson has ample historical ground for including them. They have been left in from the first. Fechner and Müller called them the 'z-cases' (*zweifelhaft*), as we have seen. Müller raised the question as to whether 'doubtful' and 'equal' judgments ought not to be separated, but concluded

[1] G. Th. Fechner, 'Elemente der Psychophysik,' 1889, I, 101ff.; 'Revision der Hauptpunkte der Psychophysik,' 1882, 24f.

[2] G. E. Müller, 'Zur Grundlegung der Psychophysik,' 1878, 28f., 33–36.

[3] Fechner, 'Revision,' 48f.

[4] Müller, 'Die Gesichtspunkte und die Tatsachen der Psychophysischen Methodik,' 1904, 104–109.

[5] E. B. Titchener summarizes the controversy: Experimental Psychology, 1905, II, ii, 278–285.

on experimental evidence that the positive impressions of 'equal' were rare, especially in trained observers, and that separate treatment of them was therefore not necessary.[1] Titchener in 1905 grouped both 'doubtful' and 'equal' cases under the heading 'u' (uncertain) or '?'.[2] Urban in 1908 had his subjects guess when in doubt which weight of the lifted pair was the heavier. Thus he obtained the categories 'heavier-guess' (hg) and 'lighter-guess' (lg), which he saw fit later to combine as 'equality' judgments.[3] In such a setting it is natural to call the region in which the 'equality' judgments are most frequent the 'interval of uncertainty.' Urban's pupils have followed his final practice. Fernberger wrote: "The equality judgment was more complex [than the other judgments] as it not only included cases of actual subjective equality . . ., but also all those cases where it was impossible for the subject to give either a lighter or a heavier judgment, usually termed doubtful cases." [4]

It is under this practice that Dr. Thomson would reject the interval of uncertainty as a measure of sensitivity because 'any subject who wishes may reduce it to zero . . . simply by determining that he will never give an answer *undecided*.' He is undoubtedly thinking of results like Fernberger's on the effect of attitude on the interval of uncertainty; but plainly he does not accept, as I think he should, Fernberger's contention that attitude can be controlled in the laboratory and that, in view of this state of affairs, it must be controlled.[5]

Fernberger's proposal was that attitude should be rendered constant by explicit instructions and observational training,

[1] Müller, 'Methodik,' 12f.

[2] Titchener, *op. cit.* II, i, 107ff; ii, 268.

[3] F. M. Urban, 'The Application of Statistical Methods to the Problems of Psychophysics,' 1908, 5f., 15, 99f., 106f., 110f., *etc.*

[4] S. W. Fernberger, 'On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli,' *Psychol. Monog.*, 1913, No. 61, 16.

[5] Fernberger, 'The Effect of Attitude of the Subject upon the Measure of Sensitivity,' *Amer. J. Psychol.*, 1914, **25**, 538–543. Fernberger refers here to another case in his own experiments and to Warner Brown's results. He has since called my attention to a study where three of eight subjects give no interval of uncertainty, presumably because of failure of attitudinal control: A. L. Ide, 'The Influence of Temperature on the Formation of Judgments in Lifted Weight Experiments,' 1919, 25 pp. (Univ. Pennsylvania thesis).

but I am going further in insisting that the instructions and experimental setting must assure a practically complete elimination of doubtful judgments. This is the point of George's study,[1] which merits, I think, considerable attention from psychophysicists. George's thesis is that in a psychophysical experiment we are dealing with a series of mental states which are a continuous function of the given series of stimuli: we vary the stimuli and note the concomitant mental variation. Under the rules of scientific experiment everything else must be kept constant including the attitude and psychophysical constitution of the subject; if they are inconstant we can no longer tell of what our judgments are a function. Moreover, if we find that any particular category is in itself an indicator of a change in attitude, we must so arrange the experiment that this form of judgment will not occur or rule it out from the results if it does occur, since the information that it yields is beside the point of the psychophysical problem, which always proposes the establishment of the dependence of judgment upon varying stimulus. George finds doubt and the doubtful judgments to be the great offenders against constancy. His method is the establishment of an *Einstellung* for constancy of attitude and the determination of (1) what categories the maintenance of the *Einstellung* rules out, (2) what categories, conversely, are noted by the observers as interfering with the *Einstellung* when they do appear, and (3) the objective evidence of attitudinal shift furnished by reaction times. George's article must be its own summary; I can not do it justice here. It is scarcely more than preliminary and his method itself may yet be called in question. It would be a slender weapon with which to combat the main body of psychophysical practice were it not for the fact that, like Dr. Thomson, many psychophysicists already know that something or other must be done to gain greater constancy of results, that the case against the doubtful judgment is plausible *a priori* [is a doubtful difference *ipso facto* less than an undoubted one?],

[1] S. S. George, 'Attitude in Relation to the Psychophysical Judgment,' *Amer. J. Psychol.*, 1917, **28**, 1–37.

and that the attitude that George proposes as an ideal has already gained some slight support from experimental usage.[1] My personal prejudice for George's conditions I shall mention presently.

2. Doubt is the most persistent offender and the one presumably most responsible for Dr. Thomson's strictures upon the interval of 'uncertainty,' but other attitudinal seducers must also be dismissed. George makes a case against expectation, a case which indicates among other things that a haphazard method of presentation is to be preferred to a serial method with partial knowledge (*e.g.*, the method of least perceptible differences).

'Reflective' judgments are the general class under which inconstancy is apt to occur. George does not find that the reflective attitude necessarily means uncontrolled variability but merely that it favors it. The reflective judgment is very often equivocally determined; the subject is in a dilemma as to his report and *decides* what to report. He may be judging on the basis of more than one criterion, and may find in a given case different categories concurrently indicated by different criteria. If he must make a univocal report for such an equivocal situation, he must make some decision, even though the decision may be for the category 'undecided.' It is the occurrence of such uncontrolled 'decisions,' I take it, that makes Dr. Thomson wish to give up the threshold as a measure of sensitivity; but his rejection of the threshold is not necessary on this score, since reflective judgments can be avoided.

George shows that doubtful judgments tend to be reflective. The doubting subject is resolving a dilemma in favor of one category or the other, and does justice to the unreported category by labelling his judgment 'doubtful.' The 'or-judgments' (*e.g.*, 'equal-or-less') are compromises in which neither side has won out. The category 'no-difference' is also often a reflective compromise. Reflection and hesitation are, however, not synonymous; a judgment may be long

[1] L. B. Hoisington, *Amer. J. Psychol.*, 1917, **28**, 588ff.; M. Kincaid, *ibid.*, 1918, **29**, 227–232; A. M. Bowman, *ibid.*, 1920, **31**, 87–90; C. C. Pratt, *ibid.*, 1920, **31**.

delayed and yet come as a simple report of mental process without any indication of resolution or of attitudinal inconstancy. In fact it seems probable that the time of formation of the judgment is one of the mental factors which under a constant attitude is a serial function of the stimulus and a feasible subject for exact psychophysical investigation.

The type of judgment furthered by ordinary psychophysical procedure is shown in the introspective analyses in Fernberger's recent monograph.[1] These descriptions he obtained under conditions analogous to Urban's except in so far as the introspection itself interfered. Compared with what I have in mind as the ideal, they show a relatively complex—often very complex—process of comparing. I should not on the basis of them expect even the degree of constancy which one actually does get. The equality judgments are typically reflective in that they involve a 'verification process,' which Fernberger, as if in support of George, seems to equate to 'doubt.'[2]

Of course what is needed to support my argument, and what is lacking, is the companion introspective study made under George's conditions. Perhaps these consciousnesses would not prove so simple as I think. My conviction that the two consciousnesses would, however, be very different affairs is dependent, I must in honesty confess, upon personal experience. I was one of George's observers and I have observed in other experiments under his conditions. I have also observed with lifted weights on the turning-top table under conditions patterned after Urban's. And the two consciousnesses are to me almost unbelievably different. Under the conventional procedure I am constantly forced into resolutions, verifications, decisions, like those that Fernberger describes, and thus I am led into doubt and discomfort, and thence into a naïve uncritical attitude which affords no assurance of rigorous constancy. Under George's conditions enough effort is required to be sure, but the mental process is kept simple or else rejected. I do not have to report a

[1] Fernberger, 'An Introspective Analysis of Comparing,' *Psychol. Monog.*, 1919, No. 117.

[2] *E.g.*, *ibid.*, p. 160.

complex situation by an inadequately simple word, nor an equivocal setting by a univocal judgment. It seems reasonable, does it not, that a report under George's conditions should be worth more than a report under Urban's, because it leads to attitudinal constancy or the detection of inconstancy when it does occur.[1]

3. There is as much to be gained by the establishment of definite serial criteria of judgment as by the elimination of unsuitable categories of judgment. In fact, these two remedies are obverse and reverse. One can not get rid of doubt, for instance, unless the judgments are based upon a univocal criterion. Two criteria that may conflict are fatal to constancy since their resolution is left to chance.

The 'stimulus error' is a term which among other things characterizes judgments where a definite mental criterion is not established but judgment is left to chance habits. The term is undoubtedly much maligned as there are cases where the direction of the attention to the stimulus is psychologically useful, as in the preliminary investigation of a new perceptual field. But in general the stimulus attitude means indefiniteness and instability of criterion, as George pointed out.[2] A recent series of studies on the various criteria that may underlie the judgments of cutaneous duality shows how fundamental to accurate psychophysical work an avoidance of the stimulus attitude is.[3] In lifted weights Friedländer's

[1] In justice to Urban it must be said that I do not believe he would experiment now as he did in 1908. *Cf., e.g.*, the 'Statistical Methods,' 1908 (*op. cit.*), with his 'Ueber einige Begriffe und Aufgaben der Psychophysik,' *Arch. f. d. ges. Psychol.*, 1913, **30**, 113–152. It is a pity that he has not yet been able to return to experimental work since writing his systematic articles. Moreover, both Urban and Fernberger have been under a special disadvantage in that they were dealing with a peculiarly refractory material, lifted weights. Isolation of a homogeneous series of univocal mental correlates of the stimulus is very difficult in the lifting process, although it has been attempted: H. Friedländer, 'Die Wahrnehmung der Schwere,' *Zeitsch. f. Psychol.*, 1920, **83**, 129–120.

[2] George, *op. cit.*, 35f.

[3] E. J. Gates, 'The Determinations of the Limens of Single and Dual Impression by the Method of Constant Stimuli,' *Amer. J. Psychol.*, 1915, **26**, 152–157; Titchener, 'Ethnological Tests of Sensation and Perception,' etc., *Proc. Amer. Philos. Soc.*, 1916, **55**, 206–215; E. deLaski, 'Perceptive Forms below the Level of the Two-point Limen,' *Amer. J. Psychol.*, 1916, **27**, 569–571; C. L. Friedline, 'Discrimination of Two Cutaneous Patterns below the Two-point Limen," *ibid.*, 1918, **29**, 400–419.

experiments, although they leave much of psychophysical accuracy to be desired, seem at least to show that different numerical results follow when attention is upon the weight and when it is upon the sensory aspects of the lifting.[1]

The more general ground for the control of criteria is the one which Fernberger took in defense of the interval of uncertainty.[2] The subjects must be constantly and effectively *eingestellt*, and the test of an effective *Einstellung* lies in preliminary trials, the taking of introspections, and the observer's full and repeated characterizations of their attitude.

4. The need for the isolation of the single judgment within the series is perhaps worth especial mention since the matter has just been [implicitly] brought to fore by Fernberger's measurement of the effect of one member of a series upon a succeeding member.[3] Under Urban's conditions with the turning-top table Fernberger found that the judgment 'lighter' of a pair of lifted weights tends to be succeeded by a judgment 'heavier,' and *vice versa*. 'Fixing' a series so that one kind or the other of sequences predominates produces as startling effects upon the form of the psychometric functions as anything that Dr. Thomson is complaining of. Plainly some sort of expectational or rhythmic effect is operative; attitude is not remaining constant. Fernberger's solution of the difficulty is to balance one sort of succession against the other, and trust that they will cancel. For myself, I could not feel secure in such a procedure; an algebraic cancellation where the factors are so little understood can not be so satisfactory as an actual elimination. The members of the series should be separated by an interval—a distracted interval if necessary—so that the intraserial effects are broken up. This course slows down the rate of experimentation and robs the turning-top table of much of its charm; but, even when relative frequencies are aimed at, I do not conceive that numbers of cases can be allowed to weigh against rigorous scientific control.

[1] Friedländer, *op. cit.*

[2] Fernberger, *Amer. J. Psychol.*, 1914, **25**, 538ff. (*op. cit.*).

[3] Fernberger, 'Interdependence of Judgments within the Series for the Method of Constant Stimuli,' *J. Exper. Psychol.*, 1920, **3**, 126–150.

5. Lest the argumental sauce obscure the meat, let me summarize. I recommend (1) that every judgment in the psychophysical experiment stand absolutely independently in its own right. One member of the series must be separated from the others in time, and by the instruction to the subject that he judge it without reference to any other member. Haphazard presentation should be the rule; at any rate serial presentations with partial knowledge should not be allowed since they connect the members and interfere with their individuality. I urge further (2) that the criteria of judgment be laid down explicitly and univocally in psychological terms. These psychological terms will be sensory in the class of experiments especially under consideration. Judgments of stimulus are often desirable, but they are not the final ideal since they are *ipso facto* equivocal.[1] The univocal character of the criteria must be tested by introspection and by the subjects' report. (3) The total *Aufgabe* under which the subject judges must be made definite in instructions, and must be more fully determined by means of repeated characterizations by the subject of his attitude and procedure. This latter check is important since much of the subject's instruction is apt to be a self-instruction. (4) The subject must be both instructed and trained to maintain a constant attitude throughout the experiment and to report lapses from this attitude. When he has learned the full meaning of this instruction, he will not give doubtful judgments nor ordinarily be doubtful, provided his task is made sufficiently easy for him by the means of the three foregoing rules. He will in like manner avoid other reflective judgments that violate the constant attitude. (5) He will probably tend to give immediate judgments, and he will be greatly helped if he is encouraged to report quickly. His times, however, will vary

[1] I have met psychologists who smile superiorly when I mention the 'stimulus-error' and even have something to say in reply about 'bigoted introspectionism,' so I know that I ought not, on purely diplomatic grounds, to bring the stimulus-error in; but unfortunately for diplomacy it belongs in. I hope some day to show that the stimulus-error is not a figment of an epistemologizing or a quibbling mind, but that it is a very real scientific devil. In the meantime let those whom numerical measures alone will impress, see Friedline, and Friedländer, *opp. citt.*

and there will occasionally be long delays [they seem like 'inhibitory jams'!] without gross shift of attitude.

Undoubtedly research will bring more means of control to the fore, but the observance of these five rules alone will, I think, give Dr. Thomson data for which the thresholds will show the degree of constancy that he desires. If in the course of doing all this he thinks that we have made the thresholds 'more physiological' and less 'moral,' well and good. It is of this sort of stuff, nevertheless, that psychology is made.

The Nature of the Psychometric Functions

In the article which has caused me to write this paper, Dr. Thomson supplies us with certain suggestions as to the nature of the psychometric functions—an exposition that employs the familiar device of the urn and balls. He gains comfort from this analogue, I take it, because within it he can show how the number of black balls necessary for a given category 'depends upon a conscious act of the subject, and can be varied, if he be so disposed, at his whim.' He gains support for his analogue (1) from the fact that it gives psychometric functions that are in accord with present statistical analysis[1] in that they are not normal curves (they are not the *phi*-function of *gamma* nor the normal bell) and (2) from the fact that the three psychometric functions are founded upon a single underlying error curve. I am quite ready to be convinced of (1), but (2) seems to me scarcely a reason, since there are other ways of founding a set of psychometric functions upon a single error function. I have suggested such a derivation elsewhere,[2] and I desire here to raise the question which of these two analogues better represents our notion of the psychophysical organism, and, furthermore, whether experiment and curve-fitting may not ultimately decide between the two and thus throw light upon the nature of sensitivity.

[1] Cf., Thomson, *op. cit.*, 304–307; 'The Criterion of Goodness of Fit of Psychophysical Curves,' *Biometrika*, 1919, **12**, 226–229.

[2] Boring, 'A Chart of the Psychometric Function,' *Amer. J. Psychol.*, 1917, **28**, 465–470.

Dr. Thomson's parable runs thus. A stimulus (or a stimulus-pair) is like an urn with black and white balls in it in a given proportion. The experimental trial is the drawing of a given (constant) number of balls from the urn. Something varies so that different numbers of black balls turn up in the successive drawings. The number of black balls drawn represents the impression; there is a series of impressions possible for a single stimulus all the way from no black balls to all black balls. (At least so I interpret Dr. Thomson.) The subject reports on this series in terms of (say) three categories. Dr. Thomson thinks that the subject decides, 'if he be so disposed, at his whim' what proportions of black balls drawn shall be reported by each of the categories, *e.g.*, whether 40 per cent. to 60 per cent. shall be reported as 'undecided' with the reports 'less' for under 40 per cent. and 'greater' for over 60 per cent. I urge that rigid criteria should replace the subject's whim. The same thing happens for the other stimuli, which are represented each by other urns in which the proportion of black balls is different. The series of urns, analogous to the series of stimuli, shows a continuously increasing (or decreasing) proportion of black balls, but the relation between the place of the urn in the series (stimulus-value) and the proportion of black balls need not be linear. Psychometric functions result that are similar to those actually obtained in practice, although they are not normal functions.

My own parable, adapted to the differential limen and the urn, separates the stimulus from the organism. A given stimulus supplies a fixed number of black and white balls, since a stimulus is ideally constant. In stimulating the organism its effect is facilitated or inhibited according to the chance disposition of the organism, that is to say, the organism is an urn from which an additional number of black and white balls is drawn to be added to the fixed number determined by the stimulus. The total resulting proportion of black balls fixes the position of the impression in the impressional series, and the report in terms of predetermined categories occurs as it does in Dr. Thomson's reasoning. The constant

tendency toward a given proportion of black balls is a function of the magnitude of the stimulus, and the variability about this constant tendency is a function of the organism (the urn). The proportion of black to white balls in the urn may be anything at all but is the same for all stimuli. The resulting ogive psychometric functions may, in a certain simple case, be normal, but there is no reason why they should be nor *a priori* presumption that they would be.

My first question is this. Is it not more reasonable to ascribe constancy to the stimulus and variability to the organism, and to assume a law of physiological variability that is fixed and independent of the magnitude of the stimulus? In other words, should not the organism be represented by a single urn with contents of fixed composition? I confess I am not clear as to what Dr. Thomson's urns symbolize nor as to precisely where in his scheme variability occurs. He has an urn of different composition for every stimulus. If variability resides in the organism, then the law of variability changes for every stimulus, and such an occurrence does not seem to me physiologically understandable.

To my second question I am not even prepared to suggest an answer. May it not be possible to determine empirically what urn-and-ball analogy best fits the facts, and thus analytically to learn something new of the laws of organic variability? A determination would be ever so much better than guesses—Dr. Thomson's or mine,—but for most psychologists curve-fitting is precarious work. If we had the proper data, would the mathematical solution be feasible? Perhaps Dr. Thomson will tell us.

THE PHYSICAL MEASUREMENT AND SPECIFICATION OF COLOR

BY LOYD A. JONES AND PRENTICE REEVES

Eastman Kodak Co., Rochester, N. Y.

A complete understanding of the subject of color involves a knowledge not only of the nature of the sensation resulting from the action of the radiant energy upon the retina, but also of the exact physical composition of that radiation. The physicist has been inclined to overlook the nature of the sensation and to regard as sufficient for a complete specification of color a determination of the exact physical composition of the radiation considered. On the other hand, the psychologist has been prone to neglect the radiation factors and to regard as sufficient a specification of the nature of the sensation resulting when this radiation acts upon the retina. It is important for the sake of continued progress in this field that both phases be given due consideration. In this paper it is proposed to present an outline of the methods whereby the various factors necessary for the complete specification of color may be determined. It is the desire of the writers to treat this subject so as to give a broad general survey of the whole field in order that the interrelations between the various factors may be emphasized rather than to deal in detail with any specific phase.

A careful consideration of the subject as a whole leads to the conclusion that for the measurement and specification of the nature of the stimulus, *i.e.*, radiant energy, analytical methods must be employed, while in dealing with the sensation the methods are necessarily of the synthetic type. The necessity of using analytical methods in dealing with the stimulus is due to the fact that radiation in general is composite in nature and must be separated into its component parts in order that each may be measured. The most generally accepted theory of radiation postulates that radiant

energy is transmitted as a form of transverse wave motion in which the wave-length or frequency may vary throughout wide limits. When radiation of certain wave-lengths reach the retina the sensation of light is produced. The wave-length is usually specified in terms of millimicrons ($\mu\mu$), the micron being one thousandth (.001) of a millimeter. The Ångstrom unit (.000,0001 mm.) is sometimes used and is one-tenth of a millimicron. The visual range is approximately from 400 to 700 $\mu\mu$ although radiation of shorter or longer wave-length may be perceived if sufficiently intense.

When any stimulus acts on the retina the resulting sensation gives no indication as to whether or not the stimulus is simple or compound in nature. This indicates that the retina is a synthetic receptor and does not recognize the individual component parts of the radiation as such but receives the mixed radiation as a single stimulus producing a single sensation. The sensation produced may be specified by two factors, brightness and color, the former being dependent on the intensity, and the latter on the quality (wave-length composition) of the stimulus. The Committee on Nomenclature and Standards of the Illuminating Engineering Society states: "Color of luminous flux is the subjective evaluation by the eye of the quality of the luminous flux. Any color can be expressed in terms of its hue and saturation." [1]

Since visual sensation is dependent upon the radiant energy emitted by some luminous source, it will be well to consider briefly the manner in which the quality of such emitted radiation may be measured and specified. This is most satisfactorily accomplished by separating the radiation into its component parts and measuring the intensity of the individual element. This factor is properly expressed in watts (or ergs per second) per square centimeter of the emitting source per unit difference in wave-length. Such values when plotted as ordinates against the various wave-lengths as abscissæ result in a graphic representation of the spectral energy distribution for the source considered. Such a curve is commonly referred to as a 'spectral energy curve' or

[1] *Trans. Illum. Eng. Soc.*, 1918, **13**, 515.

'emission curve' and when determined for all wave-lengths within the visible range constitutes a physical specification of the quality of the emitted radiation, and hence of its color. The measurement of the energy values is accomplished by use of an instrument known as the spectro-radiometer in which either a bolometer or thermopile is usually employed as a receiving element. It will not be advisable to go into a detailed discussion of such instruments and methods at the present time, and for more complete information the reader

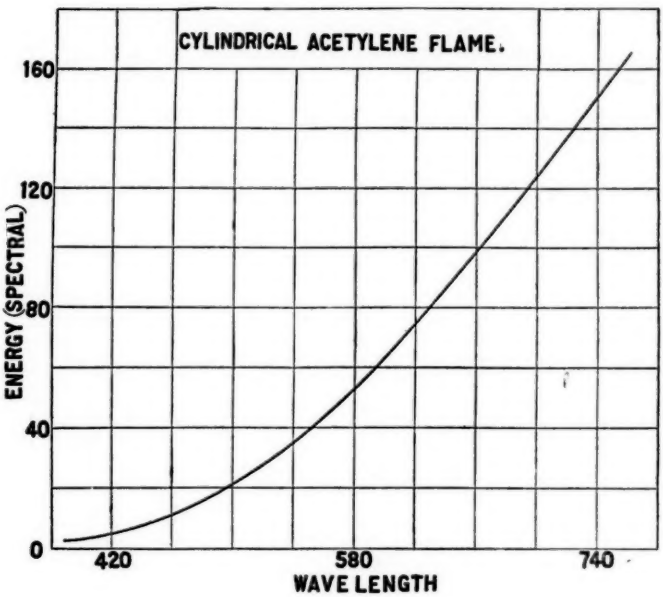

FIG. 1

is referred to the literature on spectro radiometry.[1] The spectral energy curves of many sources have been precisely determined and as a typical example, the curve of the cylindrical acetylene flame is shown in Fig. 1. A black body is probably the best standard for spectral energy distribution,

[1] Nutting, P. G., 'Outlines of Applied Optics.' Philadelphia: P. Blakiston's Sons & Co., 1912, p. 234, Chapter IX. Also see recent reports by P. D. Foote in *Trans. Amer. Inst. of Mining and Metallurgical Engineers.*

but a black body is difficult to realize in practice so that a standard acetylene burner operated under the conditions recently specified by Hyde[1] is practically identical with a black body operated at 2360° K., *i.e.*, 2087° C. Such a source is easily reproducible.

The precise determination of the emission curve of the source requires an elaborate equipment and considerable experience in manipulation on the part of the operator. However, if this function is known for one source that of any other can be determined indirectly with considerable ease by a method of spectrophotometry. The method consists simply of a comparison and quantitative measurement, wave-length by wave-length, of the intensity of the unknown source in terms of that of the known throughout the visible range. A spectrophotometer for this purpose consists essentially of an optical system such that two beams of light, one from the source whose spectral emission is known and the other from the source which is being measured, may be dispersed into their component parts. From the spectra thus formed narrow regions may be isolated and used to illuminate the parts of some suitable photometric field. In the path of one beam is situated some device such as a pair of Nicol prisms, a rotating sector, or a slit of variable width by which the intensity of that beam may be varied in a known manner. By the proper adjustment of this device, a photometric balance may be made and from the constants of the system the ratio of the intensity of the known to the unknown is determined. The measurement of this ratio at a sufficient number of points suitably spaced throughout the visual wave-length range provides the data from which the emission curve may be plotted. As examples of the most commonly used spectrophotometers may be mentioned the Lummer-Brodhun, the Brace, the Hufner, König-Martins, each of which has its peculiar advantages and disadvantages for the various special purposes in the field of spectrophotometry. Space does not permit a detailed discussion of this instrument and again the

[1] Hyde, Forsythe & Cady, *J. Frank. Inst.*, 1919, **188**, 129–130. See also Coblentz, *ibid.*, 1918, **188**, 299.

reader is referred to the literature on the subject for more complete information.[1]

Non-luminous objects are visible only by virtue of the radiant energy which they transmit, reflect, or otherwise divert, in such manner that it enters the eye and falls upon the retina. In case an object reflects or transmits to an equal extent all wave-lengths of the incident energy within the visible range, it is said to be visually non-selective or colorless. This class of objects includes all true grays, which form a scale varying only in intensity, and limited at the extremes by black and white. However, when an object transmits or reflects to an unequal extent the wave-lengths of the incident energy it is visually selective and is colored. Every known substance absorbs to some extent radiation of some wave-length and nearly all absorb very strongly at some particular wave-length or spectral region. The colors of opaque objects depend upon the ratio of reflecting to absorbing power for each wave-length.

A saturated color, that is a pure hue, reflects or transmits a very narrow region of the spectrum, *i.e.*, absorbs most of the spectrum and is, practically speaking, monochromatic. Such media are rare in nature and practice, as most objects we meet are far from being monochromatic. The spectrophotometer is used to determine the amount of transmission or reflection at each wave-length. Fig. 2 shows the curve of a green filter with wave-lengths as abscissæ and density as ordinates in one curve and transmission as ordinates in the other. This type of curve is rather typical of ordinary colored things with the maximum transmission in a certain spectral region and the somewhat gradual sloping to zero transmission on either side. A density of 1 allows ten per cent. of the incident light to pass so the corresponding transmission is 0.1, a density of 2 transmits 1 per cent. and so on. Density is the logarithm of the reciprocal of the transmission, *i.e.*, $D = \log 1/T$. If two filters are taken together their combined density is the sum of the separate densities and their

[1] Nutting, *op. cit.*, Chapter VIII. Also annual reports of Committee on Progress in Trans. Illum. Eng. Soc.

transmission is the product of the separate transmissions. The physical law of absorption stated that the absorption at any wave-length is an exponential function of the thickness of the absorbing media or in other words if a unit thickness transmits a fraction T, absorbs $(1 - T)$, then the next unit thickness will transmit the same fraction of what remains and a thickness X will transmit the fraction T^x. This law applies only to homogeneous media and monochromatic radiation.

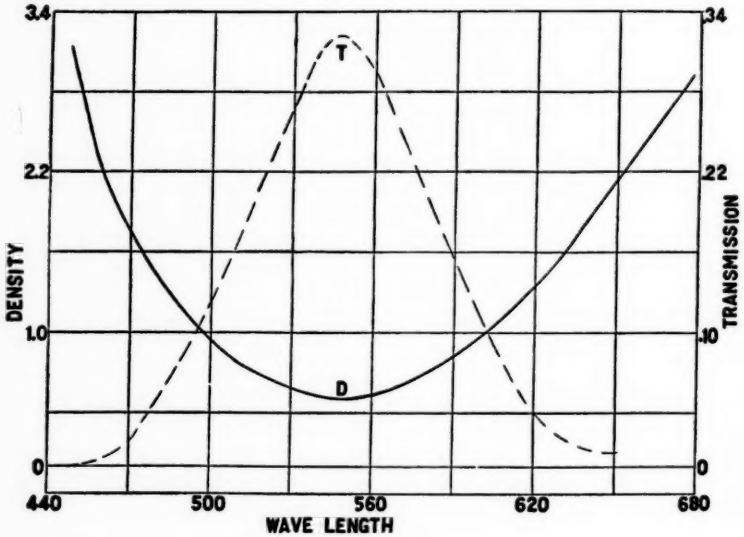

FIG. 2

A spectrophotometric curve gives the relative transmission of a filter but may be taken as indicative of color only when the composition of the incident light is known and specified as shown by the spectral energy curve. If we change the nature of the incident light through a filter, or on a reflecting surface, we also change the nature of the transmitted or reflected light. Another important thing to bear in mind is that equal energies do not produce equal brightnesses and this is illustrated in Fig. 3 which shows the so-called visibility

curve for an average eye. This curve is determined by measuring the relative amount of energy necessary at each wave-length to cause equal sensations of brightness. If we

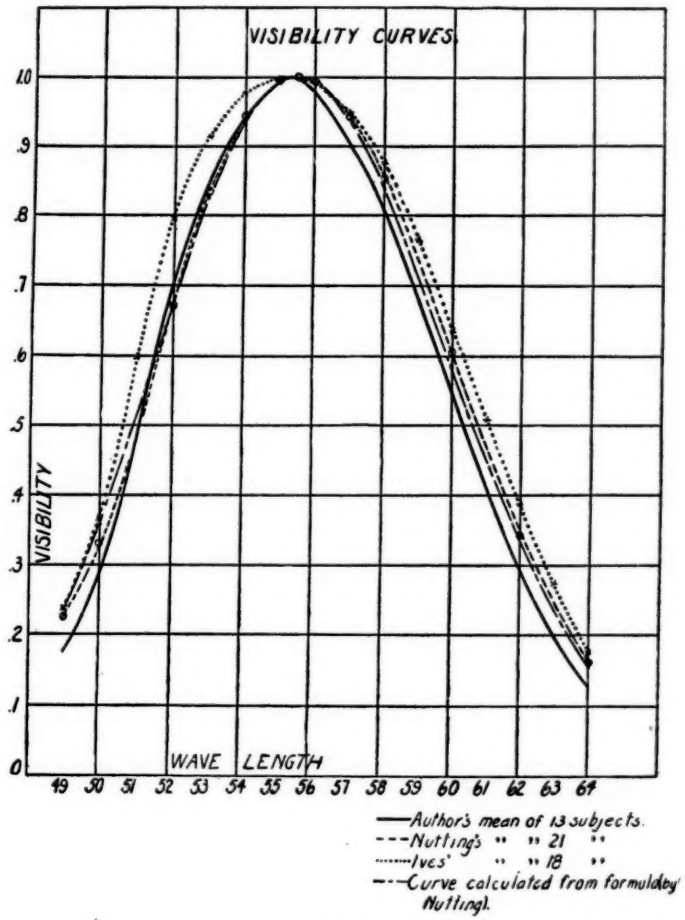

FIG. 3

take the maximum sensibility as unity (the point in the spectrum being 556 $\mu\mu$) the relative energy values for other wave-lengths in the visible spectrum are represented by the ordi-

nates of the curve.[1] This shape and position of the maximum of the visibility curve varies with intensity of illumination as we might expect from dealing with peripheral or foveal vision. The product of spectral energy and visibility at each wave-length gives luminosity and it is really the luminosity curve that we consider when treating a visual stimulus.

Spectrophotometry gives the exact physical composition of the radiation in question, *i.e.*, analyzes the stimulus, but does not provide any direct specification of the subjective factors of the color sensation resulting when this radiation impinges upon the retina. This brings us then to the synthetic method by which the specification of the sensation is made directly in terms of the subjective factors, brightness and color. Brightness is dependent upon the intensity and color upon the quality of the incident radiation. These measurements are usually made by means of a colorimeter, the action of which is based on the fact that any color can be matched by a mixture in proper proportions of monochromatic light of the correct wave-length, with white light. The wave-length of the monochromatic light used is termed the "wave-length of the dominant hue" and constitutes the specification of the hue factor. The amount of white light necessary to make a match taken as a percentage of the total mixture is known as the per cent. white and represents the saturation factor. In case the color to be matched is a non-spectral color such as a purple or a magenta the complement of the color is found and the specification stated in terms of the complementary color. Brightness is measured by a suitable photometer which may or may not be an integral part of the colorimeter itself. In the case of reflecting surfaces the intensity factor is specified in terms of reflecting power, in case of transmitting media, by the total transmission and in case of emitting sources by the intensity of the source.

In the latest colorimeters it is possible to read hue to a fraction of a wave-length in the visible spectrum and a problem which arises from such a possibility is the determina-

[1] Reeves, P., 'The Visibility of Radiation,' *Trans. Illum. Eng. Soc.*, 1918, 13, 101. The most extensive work on this subject was published by Coblentz and Emerson. Sci. Paper 303, Bureau of Standards, issued September 12, 1917.

tion of a hue scale. If we desire to measure any quantity it is first necessary to have a unit of measurement which is constant throughout the entire scale. (For example, an inch is an inch, or a pint a pint wherever it may be taken.) When examining the visible spectrum, however, we find that equal wave-length intervals do not produce equal color sensation intervals at different parts of the spectrum, so the wave-length is not a satisfactory unit in the establishment of a scale of subjective color sensation and a hue scale established with fixed wave-length intervals as a unit is therefore unsatisfactory. The question as to whether or not the difference limen is equal at all points on any sensation scale has been discussed pro and con for years but from a physical standpoint it seems permissible to make these points equal by definition and call the least perceptible difference the sensation unit and use it as such in establishing sensation scales. So the experimental problem in establishing a fundamental hue scale is to determine the relation existing between the wave-length unit and the least perceptible difference in hue for the entire range of visible radiation. Two methods of procedure present themselves,[1] one method measuring the least perceptible difference at certain intervals, say every 5 or 10 $\mu\mu$, throughout the spectrum and the other progressing step by step and measuring each least perceptible difference in hue in the scale. Fig. 4 shows the results obtained from the latter method. The greater the least perceptible difference the less the sensibility, so the sensibility can be taken as proportional to the reciprocal of the least perceptible differences. Curve A represents the hue sensibility curve obtained from the reciprocal of the limens plotted against wave-length. Curve *B* is the scale reading curve obtained by integrating the sensibility curve. Or by taking least perceptible differences as units the hue scale may be determined by direct measurement and the sensibility curve obtained by differentiating the scale reading curve. Examination of these curves shows maxima and minima at different points in the spectrum. These results agree very well with previous results obtained,[2]

[1] Jones, L. A., *J. Opt. Soc.*, 1917, **1**, 63.

[2] Steindler, C., *Sitzungsber. Wien. Acad. Wiss.*, 1906, **115**, 1.

though, of course, the final values must be the average results from a large number of observers with normal vision.

For an exact determination of the hue scale, spectral light must be used, as the development of transmitting media or reflecting surfaces up to the present time has not furnished either material with monochromatic properties. So an apparatus which will illuminate independently the two parts of a photometric field with monochromatic light of variable

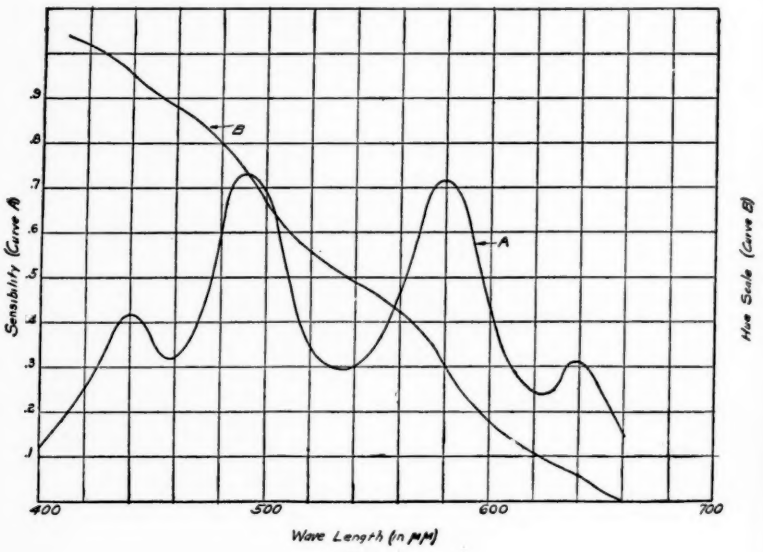

FIG. 4

quality and intensity must be used. It is quite essential that the wave-length in each part be easily varied and measured, and also important that the intensity be accurately controlled. In making a judgment of equality or difference of hue with two fields, it is important to have the intensity of those fields balanced so that an intensity difference does not influence the hue judgment. The most satisfactory apparatus used in our laboratory consisted of a Brace spectrophotometer used in connection with a Hilger spectroscope of the constant deviation type. The number of distinct

hues between 400 $\mu\mu$ and 700 $\mu\mu$ was found to be 128 when the observer started at a given point and proceeded step by step through the visible spectrum. Several more steps (about twenty) are added if we examine the non-spectral purples and magentas. Hue sensibility is nearly independent of brightness, although it is found that the sensibility is somewhat higher for brightnesses of medium value than for those of extremely high and low values.

By mixing a pure hue with white and at the same time maintaining a constant brightness, a series of colors will be

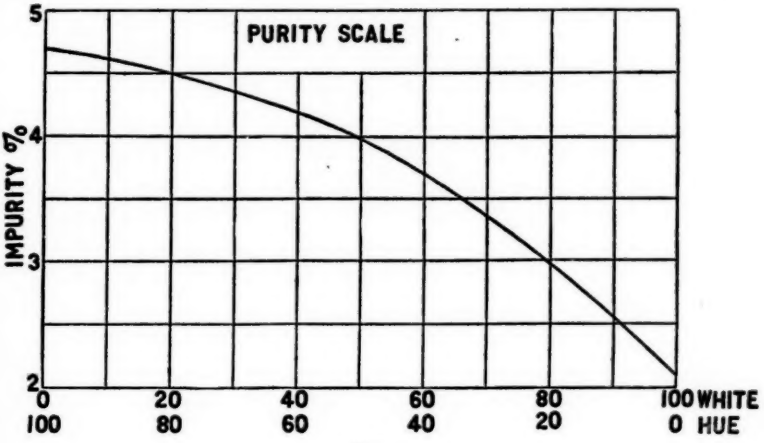

FIG. 5

obtained differing only in saturation. A specification of saturation, therefore, denotes the proximity of the color to a condition of monochromatism. Various terms such as purity, chroma, and tint are used, but the writers consider either saturation or purity the most suitable. If we start with a pure hue (monochromatic spectral light may be considered, for purposes of measurement, as having a saturation of 100 per cent.) and proceed to a saturation of zero, we find about twenty steps. Fig. 5 shows the purity scale with the per cent. white and the per cent. hue plotted as abscissæ and the least perceptible differences expressed as percentages, as ordinates.

At this point it might be well to review some of the factors

involved in determining the number of visual sensations and compare the use of spectral light and physical controls with the ordinary method of mixtures on a color wheel. In either method the number of perceptible differences we find in a hue scale or purity scale will depend on the brightness region we choose for the hue scale and the initial saturation for the purity scale. In a suitable apparatus using spectral light we are able to select very narrow regions of the spectrum and know accurately the region selected from a direct reading scale. Spectral colors are saturated. When mixing white light ("white" in this usage means any gray in color mixing) we can get a very wide range, from the full intensity of the light source by removing the Nicol prisms to total extinction when the elements are crossed. Here again the physical specifications are easily obtained from a scale reading. For the purposes of comparison it is rather easy to obtain two identical sources and, furthermore, results obtained in one laboratory will be directly comparable with those obtained in any other laboratory. With the color wheel, however, we first meet the difficulty of obtaining monochromatic discs and discs approaching a physical white or black. Then we do not know the physical specifications of the mixtures obtained from the various proportions of the discs used. Discs from the same order may differ from one another and many of them are unstable even in darkness. These variable factors not only preclude intercomparisons but also make separate results doubtful except for demonstrations. In any case where discs are used it would be advisable to include in published data the physical specifications of the discs used.

Many text books in psychology make the statement that "yellow is distinctly lighter than green; violet is darker than the other spectral hues," and consequently picture the spectral belt in the color pyramid or color spindle as being tilted upward at the yellow region and downward at the violet. This type of statement is true only for certain spectra. The prismatic spectrum of sunlight will certainly bear out these statements but any number of other spectra show a different set of conditions and it is possible to get equally bright yellows and violets. We also find statements that violet and blue

are the most saturated colors. From a physical standpoint it is possible to have any color 100 per cent. saturated and while it is true that 100 per cent. yellow is quite similar to white, so, too, other saturated colors seem to approach white when the intensity of illumination is increased. In the spectra ordinarily used the dispersion and luminosity factors easily lead one to the statements mentioned above but much more work must be done with spectral regions of equal saturations, equal brightnesses and equal luminosities before we can be positive as to the form and position of the spectral belt of the color pyramid.

The Nutting colorimeter[1] is one of the most satisfactory monochromatic analyzers, and is especially adapted to the determination of the dominant hue and per cent. white of reflecting surfaces or transmitting media. The dominant hue is read from a direct reading wave-length drum on a screw which operates a constant deviation dispersing prism. This prism controls the quality of the standard, the intensity of which is controlled by a pair of Nicol prisms and the purity by mixing white light from another source.

Another form of colorimeter is the trichromatic analyzer which is partly analytical and partly synthetic in nature. This method specifies a color by giving the relative intensities of some arbitrary red, green, and blue which when mixed together match the unknown. The red, green, and blue are obtained by the use of filters of glass, stained gelatine or other suitably colored material or by the isolation of narrow bands of the spectrum. A representative of this type of instrument is the Ives colorimeter.[2] Many other so-called colorimeters are in reality only color comparitors or tintometers and in many cases are based on arbitrarily chosen standards.

Although this paper has been a hurried review of some of the facts on color, it is hoped that the importance of accurate control of the stimulus in color experimentation has been emphasized. If results are to be duplicated in various laboratories and colored stimuli standardized some of the aforementioned facts must be observed.

[1] *Bull. Bur. Stand.*, 1913, **9**, 1; *Zsch. f. Instkund.*, 1913, **33**, 20.
[2] *J. Frank. Inst.*, 1907, **164**, 421. *Ibid.*, 1915, **180**, 673.

SUGGESTIONS LOOKING TOWARD A FUNDAMENTAL REVISION OF CURRENT STATISTICAL PROCEDURE, AS APPLIED TO TESTS

BY SIDNEY L. PRESSEY

University of Indiana

The past three or four years have been notable in psychological history for the remarkable development of statistical methods as applied to the problems of mental measurement. This advance is undoubtedly of the very greatest importance. The writer has come to feel, however, that, with the first enthusiasm in such work, there has been a tendency toward over-elaborateness and diffuseness of treatment, and a lack of directness and incisiveness in the statistical procedure. And he wishes to point out certain limitations to the present concepts of "reliability" and "validity" as applied to tests, and certain objections to the customary use of the theory of the normal curve in test building, which he feels to be of distinct importance.

The situation can most readily be made clear by a very concrete example. Suppose, then, that a high-school principal desires to give a group test for measuring general intelligence to his entering class, in order to pick out in advance those who are likely to fail in their freshman work. He has a number of scales under consideration. And he wishes evidence as to the relative merits of these scales for this purpose,—for the selection of potential failures. He will very likely be given data with regard to the comparative 'reliability' and 'validity' of these scales; information may also be produced with regard to the organization of the tests, especially in respect to the normality of the distribution of scores yielded. The present paper aims to show that no one of these three sets of facts gives that close contact, which is desirable, with the practical problem.

I. Inadequacy of the Present Concept of 'Reliability'

The principal may be urged to use a particular scale because the scale has a high 'reliability.' The exact meaning of 'reliability' must, however, first be carefully looked into. The meaning of the concept can best be understood by considering the way in which 'reliability' is usually measured. The most common method is simply to give two duplicate forms of the same test, one after the other, to the same subjects. The ratings obtained by the subjects on 'Form A' and 'Form B' are then correlated. And the closeness of the correlation indicates the reliability of the test.

The significance, and the limitations, of a measure thus obtained are fairly obvious. Two such limitations are especially important. (*a*) The measure is evidently a measure of the reliability of the sampling,—of the particular type of performance involved in the test. When one speaks of the reliability of an instrument, one naturally thinks of its reliability *for some purpose*. Such a connotation must be guarded against here. One must not come insensibly to think of the reliability coefficient of a test of intelligence, for instance, as indicating the value of the test, as a measure of intelligence. Such a conclusion is sound only if a test is a simple sampling of the ability which it is sought to measure; and this happens much more rarely than might be supposed.[1] The term 'consistency' would, therefore, seem a more accurate term; the 'reliability' coefficient indicates only the extent to which a test is consistent with itself. And it is entirely possible that a test should yield highly consistent results which were, nevertheless, not at all measures of the function which it was desired to measure.[2]

[1] It might seem, for instance, that the Courtis Scale B was a simple sampling of ability in the fundamentals. But recent research has shown the situation to be by no means so simple. (See Thorndike, E. L., and Courtis, S. A., 'Correction Formulæ for Addition Tests,' *Teachers' College Record*, 1920, **21**, 1–24.)

[2] Thus, not so many years ago, cancellation tests were frequently included in 'batteries' of tests intended for the measurement of mental endowment. (See, for instance, Pyle, 'The Examination of School Children,' Macmillan, 1913.) It now seems quite clear that cancellation tests are not good tests of intelligence. (See McCall, 'Some Correlations between Mental Traits,' Teachers' College, 1916.) But cancellation tests appear to be quite 'reliable' measures,—they are simply not good tests of general intelligence. They are, therefore, not 'reliable' for the purpose for which Pyle used them.

It must also be kept in mind (*b*) that such a measure of the reliability of the sampling may be considered an adequate measure for this purpose only if the scores obtained on 'Form A' and 'Form B' may be considered entirely random samplings of performance on such a test. Usually they cannot be so considered. There may be an initial difficulty with directions at the beginning of 'Form A' and a slight fatigue toward the last of 'Form B.' What is, with many of the tests, more important—the method as described above tells us nothing whatever about the 'consistency' of the results from one examiner to another, one scorer to another, from one day to another, or one time of the day to another.

To come back to the original problem, then: such a measure of the consistency of the test with itself, under certain circumstances, tells the high school principal surprisingly little as to the value which that test may have in distinguishing his potential failures from the rest of their class.[1] And information with regard to the 'validity' of the scale is naturally turned to, to settle this practical question.

II. The Artificial Nature of Current Concepts Regarding Validity

The principal is, then, urged to use a particular scale because the scale has a high 'validity' as a measure of general ability. That is, data are presented showing that the scale gives results having a high correlation with independent criteria as to general intelligence, and congruence with current theories regarding the nature of general intelligence,—there

[1] The writer is inclined to feel that most problems of consistency can best be dealt with in general terms. That is, what difference, in general, may one expect in test results if one tests Monday instead of Friday, at 9 o'clock instead of 3 o'clock? What difference, in general, may be expected, with a given type of directions, from one examiner to another? What differences, with various scoring methods, may be expected from one scorer to another? What differences may result in the score of an individual as the result of fluctuations from one time to another, in general feeling tone, energy, vigor, health? The writer believes that, until evidence to the contrary appears, it may be taken for granted that such factors affect all tests in more or less the same way; certain general theorems with regard to their operation should, then, be possible—or general precautions taken. The only problem of consistency that needs specific determination for each test would then be consistency as it relates to the subject matter of the test.

is a regular rise in score from year to year until maturity, a relative freedom from the influence of specific training, and so on. This concept of 'validity' is also, the writer feels, beside the point, if not misleading, so far as the practical problem of the high-school principal is concerned. And again there are two difficulties.

In the first place (*a*), since the extent to which general intelligence is the fundamental factor, in conditioning success and failure in the Freshman year of high school, is not known, the usefulness of the scale (even if proven a satisfactory measure of general intelligence) is still an unknown quantity. Stability of character, willingness to apply oneself even though the restraints of grammar school supervision are now removed, interest in the more mature subjects of the high-school curriculum—such elements are probably more important than is often supposed, in the total situation.[1] Differences in the adequacy of previous preparation may also be of importance. So proof of the 'validity' of a scale as a measure of 'general intelligence' is by no means proof of the value of the scale in sorting out potential failures among these high-school freshmen. In fact, it might almost be said that in proportion as the scale measured one element only, in a complex situation, to just that extent was the scale inadequate for dealing with that total situation![2]

It remains to be pointed out, however, that even though

[1] For a discussion of this tendency to overestimate the comparative importance of intelligence see Rosenow, Curt, 'Is Lack of Intelligence the Chief Cause of Delinquency?' PSYCHOL. REV., March, 1920.

[2] The more extreme theories in regard to general intelligence surely make up, in the aggregate, an extraordinary concept. It should surely be kept in mind that it is, in the first place, an analytical concept and so dependent for its character upon the methods of analysis employed. It should also be pointed out that such a concept naturally receives successive accretions in the way of theory and may, by a mental synthesis largely adventitious to the facts, acquire a reality which is very largely an artifact. Scores on various tests are lumped and the aggregate used as a measure of general ability. A more or less close relationship is naturally found between such an aggregate and the average marks of the children in school. Teachers, especially in the grades, naturally think of the child's work as a whole, and give marks showing high correlations between abilities in different subjects. And the children come to this attitude and react to their school work as a whole. And—the whole situation is cumulative. One might, in fact, imagine the concept of general ability thus developing even though abilties were, as a matter of fact, diverse and uncorrelated.

the demonstration of a close correlation between the scale in question and a general intelligence *were* supplemented by evidence that general intelligence was the fundamental factor in the situation, still the suitableness of the scale for the particular problem would remain to be shown. That is, (*b*) the usual method for stating relationships between two variables—the correlation coefficient—does not express satisfactorily the nature of that relationship, for diagnostic purposes, at a particular point in the distribution. The problem is: How unmistakably will the scale set off the lower 15 per cent. or so in scholastic ability? A correlation coefficient is only very general evidence in regard to this particular matter.[1] And it is evidence with regard to such diagnostic efficiency that the school principal should require.

III. The Irrelevancy of the Theory of the Normal Curve in Practical Problems in Classification

Proof of the validity of a scale as a measure of general intelligence is, then, not proof of the value of that scale for sorting out potential high-school failures, since failure is not conditioned by general ability alone, and since the diagnostic efficiency of a measuring instrument is not the same thing as the general relationship of that instrument to the factor concerned. A third set of facts may, nevertheless be introduced in evidence of the value of the scale in question. It may be pointed out that the tests of the scale are very carefully constructed so that equal increments on the scale represent equal increments in ability, and so that the total distribution of abilities yielded is closely similar to the distribution of abilities that would be expected according to

[1] See, for instance, Thurstone, L. L., 'Mental Tests for College Entrance,' *J. of Educ. Psychol.*, March, 1919 and Pressey, S. L., 'Suggestions with Regard to Prof. Thurstone's "Method of Critical Scores,"' *J. of Educ. Psychol.*, December, 1919.

The writer has often wondered whether the early introduction of the Pearson products-moments formula for calculating the correlation coefficient has not hindered rather than helped the study of relationships, in psychology. There are, of course, no right and wrong methods; methods are simply more or less adequate to the data and the problem in hand. One could almost say, dogmatically, that the particular type of data and problem to which the Pearson method is applicable were relatively rare. Most practical problems require a two or threefold division.

the theory of the 'normal curve.' Once more the writer would object to the relevancy of the information to the practical problem, and on two counts.

(*a*) Construction of the scale so that equal increments of ability are related to equal increments in score means, probably, transmutation of values in terms of the per cent. passing different items into positions on the normal curve or some such procedure.[1] It need only be said here that items which give a satisfactory scaling on such a curve need by no means be the most diagnostic items. An item may appear in a test because it is the only item appearing at 1.52 P.E. (when scaled as mentioned above) or it may appear in a test because most of the potential failures cannot pass it and most potential successes can. The last criterion is obviously the fundamental one if the problem in hand is to obtain a test that shall most completely differentiate the potential failures.

(*b*) It may also be pointed out shortly that for the particular practical problem under consideration a normal distribution of scores is hardly to be desired. If a scale sets off the potential failures very completely, it will lump the assured failures at the bottom and the assured successes at the top, and spread out the questionable cases in between. In short, equal increments of ability and a normal distribution of scores are *not* to be desired if the greatest efficiency, for the practical problem postulated, is sought.

IV. Discussion

Well—most of these points seem obvious enough, perhaps. But the concept back of them indicates a fundamentally different statistical attack, in the development and use of tests. If differentiation of the potential failures in high school is an important problem, why not build a scale specifically for that purpose? Select items simply according to their ability to make the desired division. Combine those items so that such a lumping of cases at the two extremes is obtained; the reverse of the normal distribution is the distribu-

[1] Of which procedures, transmutation of percents passing at different chronological ages into supposed units of mental growth is surely more questionable still.

tion to be desired.[1] Then measure the value of the test by measuring its 'efficiency' in dealing with the practical problem for which the scale has been designed. Deal with each important problem in some such empirical and concrete fashion. And, if, out of a large number of such attempts, there emerge certain unitary factors,—a general ability, a series of character types, or what not,—well and good. But the postulation of such elements in advance, with verification primarily by reference back to these postulates, is both an unscientific and a practically dangerous proceeding.

First a very specific problem; then, after that—*everything subservient to the solution of it!* Every item chosen with reference to that one problem, every method aiming only at the most direct and empirical solution of that problem—no hypotheses, as thoroughly empirical treatment as may be! The result will be, the writer believes, an essentially new statistical approach (methods now in use suggest something of this sort, particularly the methods used in the development of the army trade tests). Such a revision of methods is, the writer has come to feel, necessary, for a clarifying of the total situation.[2]

[1] Is this not really the solution of the problem of the normal curve in mental measurement? (See, for instance, Boring, *Amer. J. of Psychol.*, January, 1920). The actual distribution of various traits is a matter of academic interest only. But meantime, the distribution to be sought in test work will be determined by the problem.

[2] And now the apology! There is little essentially new in the paper, of course. (In fact, it should be said that a detailed discussion, with full use of the literature, was first attempted, but was found to extend beyond reasonable limits.) The important thing, however, is the total implication of the various points presented. Our statistical methods as applied to tests have been largely borrowed methods,—and methods borrowed from the descriptive sciences. So the question has been: What is the test measuring, and how accurately is this thing being measured? But mental testing is not a descriptive, but a technical science. And the question should be, instead: What are we trying to do, and how well are we doing it? The distinction is, the writer believes, of the very most fundamental importance, involving fundamental differences in statistical approach.

It remains to be mentioned that the points made apply equally to measures of achievement in the school subjects or other like tests. Instead of measuring "ability in arithmetic" in the eighth grade—and then commenting mildly on the extent to which arithmetical ability in the eighth grade overlaps on the seventh, why not tackle a definite practical problem,—attempt to define the passing point in arithmetic for the eighth grade? The distribution, again, should be bi-modal, not normal,—and the other points mentioned follow.